A discrete X-ray transform for chromotomographic hyperspectral imaging

THESIS

David J. Cooke, 2d Lt, USAF

AFIT-ENC-13-M-05

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government.

AFIT-ENC-13-M-05

A discrete X-ray transform for chromotomographic hyperspectral imaging

THESIS

Presented to the Faculty

Department of Mathematics and Statistics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science

David J. Cooke, 2d Lt, USAF, BS

March 2013

AFIT-ENC-13-M-05

A discrete X-ray transform for chromotomographic hyperspectral imaging

David J. Cooke, 2d Lt, USAF, BS

Approved:

Dr. Matthew Fickus (Chairman)	Date
Lt Col Brian McBee (Member)	Date
Capt Dustin G. Mixon (Member)	Date

AFIT-ENC-13-M-05

Abstract

The United States Air Force has a pressing need for new methods of hyperspectral imaging. All current hyperspectral imaging technologies require long exposure times, since each involves filtering the available light, either spatially or according to color. We consider a recently proposed method for hypserspectral imaging that promises shorter exposure times. This new method applies the mathematical principles of tomography to the hyperspectral data cube. Known as chromotomography, this method uses a spinning prism to essentially capture the integrals of this cube over many rotations of a single line. This thesis addresses some of the mathematical issues that arise when trying to reconstruct a hyperspectral image from chromotomographic measurements. After reviewing some of the mathematical shortcomings of the current state of the art—which arise from the technical difficulties of working with the continuous-variable X-ray transform—we make three contributions. First, we introduce a mathematically rigorous, discrete, X-ray transform that is somewhat faithful to its continuous cousin. Second, we show how under a few simplifying assumptions, our discrete transform can be generalized so as to provide a good approximation of the continuous one. This discretization allows us to apply modern finite-dimensional optimization methods to the chromotomographic reconstruction problem. Our third contribution is to apply a popular new example of such a method, known as Split Bregman iteration.

Acknowledgements

David J. Cooke, 2d Lt, USAF

Table of Contents

Page

A discrete X-ray transform for chromotomographic hyperspectral imaging

I. Introduction

The United States Air Force has a pressing need for new sensor technologies that give us the capability to analyze the chemical nature of spontaneous and brief real world events [13, 17]. A popular modern method for remotely sensing chemical information is *hyperspectral imaging*, that is, taking photographs with many color channels, unlike traditional photographs that either have one color channel (grayscale) or three color channels (red-green-blue). Unfortunately, all current hyperspectral imaging technologies require long exposure times, since each involves filtering the available light, either spatially or according to color. That is, even state-of-the-art hyperspectral sensors are incapable of imaging *transient* events [3]. For this reason, in the past decade, researchers have been proposing new hyperspectral imaging systems that make better use of the available light, thereby allowing shorter exposure times [1, 3, 6–8, 10, 13, 18].

Some of these new methods of hyperspectral imaging are based on principles of *tomography*. Traditional tomography passes waves of energy through solid objects (such as human bodies) to construct three-dimensional images (data cubes) representing the objects' composition. Some recently proposed methods of hyperspectral imaging involve tomography with color. To be precise *chromotomography* (CT)—not to be confused with chromatography—is a type of hyperspectral imaging that uses the mathematical principles of tomography in order to capture a hyperspectral image via the use of a spinning prism [6]. This thesis addresses some of the mathematical issues that arise when trying to reconstruct a hyperspectral image from the data produced by a CT camera such as the chromotomographic experiment (CTEx) developed at the Air Force Institute of Technology (AFIT) [17].

In particular, we make three major contributions: (1) a mathematically rigorous, discrete model of a CT camera system that is somewhat realistic; (2) an extension of this theory which shows how to slightly generalize our discrete transform so as to make it approximate our continuous one very well; and (3) the first application of a popular new numerical optimization technique—the *Split Bregman method*—to our reconstruction problem.

1.1 Mathematical preliminaries

We now briefly outline our approach. We consider a linear model, that is, we treat the camera system as a linear operator L. This enables us to examine the reconstruction problem using *linear least squares*. In order to do this we need an expression for the operator itself, as well as one for its adjoint (transpose) L^*. We rediscover the previously known fact that their composition, L^*L is a filter, meaning that it is diagonalized by the Fourier transform. Using this fact, we find that we are dealing with an operator with a gigantic null space; in the literature, this space is known as the *cone of missing information*. This means that we are not able to directly reconstruct the original image from the CT camera measurements. That is, there is information that we lose in the system, and we cannot get it back. To bypass this issue, we now realize that we are not exploiting everything we know about our circumstances: we want to reconstruct images of *natural* scenes, and such images typically are *sparse* or have sparse gradients. This motivates us to investigate the most recent research on how to solve linear systems with large null spaces subject to the constraint that the solution is sparse or has a sparse gradient. We now consider each of these points in greater detail.

A hyperspectral image can be thought of as a continuous three-dimensional data cube, namely as a function, $f(x, y, z)$ where x and y are the spatial coordinates of the image and z is the frequency of the light. Many current hyperspectral imaging systems fall into two types. In the first type, different color filters are placed in front of the camera. Each picture taken through one of these filters corresponds

to a single z-cross-section of $f(x, y, z)$. That is, such a system measures $f(x, y, z)$, one horizontal cross-section at a time. A second type of system uses the *push broom* sensor method, which passes the available light (of all frequencies) through a vertical slit and then through a horizontally-aligned prism. The light that passes through the slit corresponds to a single x-cross-section of $f(x, y, z)$. When passed through a prism, this two-dimensional cross-section (one spatial dimension, one frequency dimension) is transformed into an intensity image with two spatial dimensions which is then captured with a conventional (grayscale) camera. This enables us to, once again, measure the original data cube one slice at a time; here, x is fixed while y and z vary.

Unfortunately, both of these methods have problems with capturing transient events, that is, events that occur over a small interval of time. With the color-filter method, we may see the event, but will only capture it in a few color bands. With the push broom method, we may not see the event at all, and even if we do, we will only measure a small vertical slice of it.

In this thesis, we investigate a modified version of the push broom system in which the slit is removed, that is, all of the available light is allowed into the camera, and is then dispersed by the prism. Though this allows more information to enter the camera, it has the unfortunate side-effect of producing a horizontal blurring effect. We can model such a system with the following linear operator:

$$(Lf)(x, y) = \int_{-\infty}^{\infty} f(x - z, y, z)\, \mathrm{d}z. \tag{1}$$

To understand this formula, consider a very simple hyperspectral image that only has three color channels: red, green, and blue with frequencies -1, 0, and 1, respectively. The operator L acts as an adding procedure that sums the number of photons that accumulate at pixel location (x, y) in the intensity image produced by the camera. In this case, $(Lf)(x, y)$ represents the number of green photons at the hyperspectral

image's (x, y) coordinate, along with the number of red photons at $(x + 1, y)$ and the number of blue photons at $(x - 1, y)$. This corresponds to the actual number of photons (of any color) measured by our grayscale camera at pixel location (x, y), since our ideal prism allows green photons to pass straight through, but shifts red and blue photons one unit to the left and right, respectively.

Note that in (1), we are assuming that the prism is oriented horizontally, namely at a $0°$ rotation angle. If we were to instead rotate the prism by $90°$ counterclockwise, we would suffer a vertical blurring, and we would instead be integrating $f(x, y - z, z)$ over all frequencies z.

This blurring—a consequence of our desire to allow more light into the camera—also unfortunately leads to a loss of information. In an attempt to recover this missing information, we use a system that spins the prism. That is, for any possible prism angle θ, we produce an intensity image corresponding to a "projection" of the original hyperspectral image along that axis. Mathematically, we model this process as a generalization of (1). In particular, we integrate over every possible xy-rotation of the lines we integrate over in (1). Thus, the measured data we produce is of the form $(Lf)(x, y, \theta)$, where x and y are spatial coordinates and θ is the prism's rotation angle. To be precise, we now define Lf as

$$(Lf)(x, y, \theta) = \int_{-\infty}^{\infty} f(x - z\cos\theta, y - z\sin\theta, z)\,\mathrm{d}z. \tag{2}$$

In the literature, such an angle-indexed collection of line integrals is known as a continuous *X-ray transform*. Such transforms have long been studied in the field of tomography. Given a data cube $f(x, y, z)$ we model our camera measurements as

$$g(x, y, \theta) = (Lf)(x, y, \theta) + \varepsilon(x, y, \theta). \tag{3}$$

Whereas f is a real-valued function over $\mathbb{R}^3$, g is a real-valued function over $\mathbb{R}^2 \times \mathbb{T}$, where $\mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$ is the set of all possible prism angles. In order to build a useful CT camera, we need a mathematical method for reconstructing f from g. That is, if possible, we would like to invert the X-ray transform. Unfortunately, as we now discuss, this transform is not invertible.

Indeed, note that due to the noise term in (3), the equation $Lf = g$ may not even have a solution. The standard way to get around this issue is to instead consider the linear least squares version of this equation. That is, we find the f that makes Lf as close to g as possible in the L^2 sense:

$$\underset{f \in L^2(\mathbb{R}^3)}{\arg\min} \|Lf - g\|_2^2 = \underset{f \in L^2(\mathbb{R}^3)}{\arg\min} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |(Lf)(x, y, \theta) - g(x, y, \theta)|^2 \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z. \quad (4)$$

To be clear, the arg min operator returns the set of f's for which $\|Lf - g\|_2^2$ attains its minimum value. This is not to be confused with the min operator which, given the same argument, would return the minimum value of $\|Lf - g\|_2^2$ over all f. Here, we are making the (reasonable) assumption that $f \in L^2(\mathbb{R}^3)$, $g \in L^2(\mathbb{R}^2 \times \mathbb{T})$ as well as the (unreasonable, probably false) assumption that the operator L given in (2) is a well-defined bounded linear operator between these spaces. It is this lack of rigor that motivates our finite-dimensional CT model given in Chapters III and IV; until then, we shall wave our hands. If L was a bounded linear operator between Hilbert spaces, the standard Hilbert space theory of linear least squares gives that the solutions to the minimization problem in (4) are precisely the solutions to the *normal equations*

$$L^*Lf = L^*g, \quad (5)$$

where $L^* : L^2(\mathbb{R}^3) \to L^2(\mathbb{R}^2 \times \mathbb{T})$ is the *adjoint* of L, namely the bounded linear operator that satisfies $\langle f, L^*g \rangle = \langle Lf, g \rangle$. In the next chapter we informally show

that L^* has the form

$$(L^*g)(x,y,z) = \int_0^{2\pi} g(x + z\cos\theta, y + z\sin\theta, \theta)\,\mathrm{d}\theta. \tag{6}$$

In the tomography literature, L^* is known as *back projection*. To illustrate how L^* behaves, first consider how the operator L treats a single frequency of light z_0 emitted from a specific spatial coordinate (x_0, y_0). Recall the intensity of light at this point is denoted $f(x_0, y_0, z_0)$. Under the action of L, the value of f plays a role in several values of $(Lf)(x, y, \theta)$. In particular, if f were a Dirac-δ function based at (x_0, y_0, z_0), the nonzero values of Lf would trace out a helix in $\mathbb{R}^2 \times \mathbb{T}$. This helix is "centered" at $(x_0, y_0) \in \mathbb{R}^2$ and has radius z_0. In particular, if we were to shine a fixed laser into our camera, the rotation of the prism would cause this point to trace out a circle in the xy-plane over time. Looking at (6), we see that the adjoint of L integrates g over this helix in order to form the value $(L^*g)(x_0, y_0, z_0)$. In short, $(L^*g)(x_0, y_0, z_0)$ sums those values of $g(x, y, \theta)$ which, if g were of the form Lf, depend on the value $f(x_0, y_0, z_0)$. In previous research, L^* has been described as *shifting-and-adding* the camera measurements [17].

Having the expressions for L and L^* given in (2) and (6), respectively, we return to the normal equations (5). The right-hand side of (5) is determined by back-projecting (shifting-and-adding) the camera measurements g. Meanwhile, the left-hand side of (5) is determined by the composition of L and L^*. In the next chapter, we informally show that the operator L^*L is of the form

$$(L^*Lf)(x,y,z) = \int_0^{2\pi}\int_{-\infty}^{\infty} f(x - t\cos\theta, y - t\sin\theta, z + t)\,\mathrm{d}t\,\mathrm{d}\theta. \tag{7}$$

From (7), we can see that L^*L integrates the values of f that lie on the surface of a cone. To clarify, note that in (7) we integrate over all points of the form $(x, y, z) - t(\cos\theta, \sin\theta, -1)$. For any θ, $(\cos\theta, \sin\theta, -1)$ is a point on the unit circle

in the $z = -1$ plane. As t varies, the values $t(\cos\theta, \sin\theta, -1)$ trace out a straight line passing through this point and the origin. As θ varies, these lines trace out a cone whose vertex lies at the origin, and has "slope" 1 with respect to the xy-plane. Subtracting the values $t(\cos\theta, \sin\theta, -1)$ from (x, y, z) essentially translates this cone so that its vertex lies at (x, y, z). That is, L^*L performs a *rolling average* of the values of f over this cone's surface. In the image and signal processing literature, such "rolling average" operators are known as *filters*. Moreover, it is well-known that filters are best understood in terms of the (unitary) *Fourier transform*, namely the operator, $E^* : L^2(\mathbb{R}^3) \to L^2(\mathbb{R}^3)$, defined as

$$(E^*f)(\alpha, \beta, \gamma) := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{e}^{-2\pi \mathrm{i}(\alpha x + \beta y + \gamma z)} f(x, y, z) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z.$$

In particular, in the next chapter, we will show that in the Fourier domain, L^*L becomes a pointwise multiplication operator:

$$(E^*L^*Lf)(\alpha, \beta, \gamma) = (E^*f)(\alpha, \beta, \gamma) \times \begin{cases} \frac{2}{\sqrt{\alpha^2+\beta^2-\gamma^2}}, & \gamma^2 < \alpha^2 + \beta^2, \\ 0, & \gamma^2 \geq \alpha^2 + \beta^2. \end{cases} \tag{8}$$

From (8), we can see that the operator L^*L destroys information, and we cannot get it back. In particular, in the Fourier domain, L^*L multiplies the values of E^*f by positive scalars if $\gamma^2 < \alpha^2 + \beta^2$ and by zero if $\gamma^2 \geq \alpha^2 + \beta^2$. That is, the null space of L^*L corresponds to those functions $f \in L^2(\mathbb{R}^3)$ whose Fourier transforms are completely supported on the cone

$$\{(\alpha, \beta, \gamma) : \gamma^2 \geq \alpha^2 + \beta^2\}. \tag{9}$$

Moreover, since $N(L^*L) = N(L)$, we can see that our camera system is not one-to-one: two hypercubes f_1 and f_2 can yield the same camera measurements even if their Fourier transforms are different, provided these differences only occur at frequencies

that lie within the cone (9). In the tomography literature, this null space is known as the cone of missing information.

Since L's null space is so large, we are forced to make additional assumptions about f in order to have any chance of reconstructing it from $g = Lf + \varepsilon$. The purpose of our camera system is to observe transient events occurring in natural scenes, meaning we may want to restrict ourselves to a class of f's that are either sparse or have sparse gradients. To be precise, we say that a vector is *sparse* if it has very few non-zero entries. An example of a natural scene that is sparse would be a clear night sky (a nearly black background with a few stars), or, alternatively, the fireball of an explosion seen at night. That is, a sparse image is one that is mostly black with only a few spots of bright color. On the other hand, if the gradient of f is sparse this means the image can be broken up into large regions where $\nabla f = 0$, that is, regions in which the color is constant. Some might describe these scenes as "cartoon-like" because their appearance is similar to paint by number.

Mathematically, sparsity relates to a zero "norm," which counts the number of non-zero entries in a vector. That is, in order to find the sparsest solution to the normal equations, we would like to solve

$$\arg\min_f \|f\|_0 \text{ subject to } L^*Lf = L^*g. \tag{10}$$

Unfortunately the zero "norm" is not a norm: it does not distribute over scalar multiplication. More importantly, optimization is mostly calculus-based, meaning that in order to solve an optimization problem such as (10), we would like to use a norm that is differentiable. The zero "norm" is not differentiable, suggesting that we should use a different norm to solve our problem. Recently, researchers have begun to use the 1-norm as a proxy for the zero "norm;" the 1-norm of f is defined as

$$\|f\|_1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |f(x,y,z)| \,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z.$$

Though not differentiable everywhere, this norm is nevertheless a convex function of f meaning it can be analyzed with the classical convexity-based generalizations of calculus-based optimization. Therefore, instead of solving (10) we try to solve

$$\arg\min_f \|f\|_1 \text{ subject to } L^*Lf = L^*g. \tag{11}$$

It turns out that, even with this relaxation, finding a solution to (11) is too difficult mathematically because of the size of L's null space. To account for this problem, we apply a standard trick, namely we *regularize* (11) by adding a term that penalizes large values of $\|Lf - g\|_2^2$. That is, to find a sparse f for which Lf is close to g, we will solve

$$\arg\min_f \|f\|_1 + \frac{\lambda}{2}\|Lf - g\|_2^2, \tag{12}$$

where λ is some experimentally chosen weight. On the other hand, to find an f with a sparse gradient we solve

$$\arg\min_f \|\nabla f\|_1 + \frac{\lambda}{2}\|Lf - g\|_2^2. \tag{13}$$

In the literature $\|\nabla f\|_1$ is sometimes known as the *total variation* norm of f.

In practice, we would attempt to solve (12) or (13) for larger and larger values of λ and use the, for lack of a better term, "eye test" to find the best solution. Notice though, that as $\lambda \to \infty$ in (12) and (13), the 1-norm term becomes insignificant and we are essentially again solving the ill-conditioned problem of minimizing $\|Lf - g\|_2^2$. The *Bregman method* is a recently introduced, yet already very popular, method for bypassing this issue. That is, it provides an alternative, more numerically stable approach for minimizing $\|f\|_1$ or $\|\nabla f\|_1$ subject to the constraint that $L^*Lf = L^*g$. We in particular apply the most recent variant of this algorithm, known as the *Split Bregman method*. This is an iterative process in which we "split" our optimization

problem into simpler ones that are easier to solve, making use of a technique known as *shrinkage* (*soft-thresholding*).

1.2 Outline

This introduction has served to build up our intuition about the CT reconstruction problem, as well as highlight some of the key quantities relating to it. In Chapter II, we delve deeper into the continuous X-ray transform $(Lf)(x, y, \theta)$, and informally derive the important equations given above. Then, in Chapter III, we will reinvent the X-ray transform in the discrete setting, which allows us to replace the hand waving of Chapter II with a rigorous analysis. Whereas the continuous model we present in Chapter II makes intuitive sense but is not rigorous, the discrete model in Chapter III is completely rigorous, but is less faithful to the real-world physics of our camera system. In Chapter IV, we refine this discrete model so as to make it more physically realistic. In Chapter V, we use the Split Bregman method along with the ideas of Chapters III and IV to perform CT reconstruction experiments.

1.3 Major Contributions

For the past twenty years, several teams of researchers have applied principles from tomography in order to attempt to build better hyperspectral imagers [1, 3, 6–8, 10, 13, 18]. Some of this previous work inspired our derivations involving the operator L given in Chapter II. Previous research has also made some of the same connections that we make: recognizing the problem is tomographic in nature, and using Fourier analysis to study it [1, 8].

However, much of the mathematics that has been previously applied to the CT reconstruction problem is not rigorous because it is difficult to prove results in infinite-dimensional spaces. In short, the integral operators that have been previously applied to this problem are not well-defined, bounded linear operators between prescribed vector spaces. We take the novel approach of attacking this problem from

a purely discrete perspective, making all of the mathematical derivations rigorous and correct. That is, we are the first to apply recently introduced principles of discrete tomography [2] to the CT reconstruction problem. This is the first major contribution of this work.

The second major contribution builds on the results of our first contribution. Making only a few simplifying assumptions—namely that our continuous function f is periodic in the spatial and frequency domains and that it has been sampled so finely that it is effectively piecewise constant—we show that the continuous X-ray transform boils down to a slight generalization of our aforementioned discrete model. Realizing this, we then use the theory stemming from our first contribution to gain a more realistic understanding of a CT camera system.

Our third major contribution is that we are the first to apply a recently introduced, state-of-the-art numerical optimization technique—the Split Bregman method—to the CT reconstruction problem. Over the past few years, this method has become very popular in the sparse image processing community. However, due to its newness and mathematical sophistication, there are still many reconstruction problems to which it has not yet been applied. New fast reconstruction algorithms such as the Split Bregman method are needed in order for CT research to move forward, since only then will researchers be able to perform the numerical experimentation and simulation needed to address important, basic questions about their camera systems. That is, reconstruction algorithms such as the Split Bregman method allow us to focus on the real issues with our CT system, and what kinds of improvements are necessary in order to make the system function better.

II. A continuous mathematical model of chromotomography

In this chapter, we continue to discuss the continuous X-ray transform, namely our operator L, as defined in (2). In particular, in this chapter, we provide the important derivations concerning L that we glossed over in the introduction. As we shall point out, these derivations are not completely rigorous due to the mathematical subtleties of infinite-dimensional spaces. Nevertheless, these derivations are on the right track. In the coming chapters, we will generalize the techniques introduced here to the discrete, finite-dimensional setting. There, they become completely rigorous. In order to use the Hilbert space-based theory of linear least squares, we assume our continuous hyperspectral image f lies in the Hilbert space

$$L^2(\mathbb{R}^3) = \left\{ f : \mathbb{R}^3 \to \mathbb{R} \;\middle|\; \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f(x,y,z)|^2 \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \, < \infty \right\}.$$

The inner product on this space is defined as

$$\langle f_1, f_2 \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [f_1(x,y,z)]^* \, f_2(x,y,z) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z,$$

where α^* denotes the complex conjugate of any real or complex scalar α. As with any Hilbert space, this inner product defines a norm: $\|f\| := \sqrt{\langle f, f \rangle}$. Now recall the operator L defined in (2), which integrates f over a set of lines in $\mathbb{R}^3$:

$$(Lf)(x,y,\theta) = \int_{-\infty}^{\infty} f(x - z\cos\theta, y - z\sin\theta, z) \, \mathrm{d}z.$$

Since f lies in $L^2(\mathbb{R}^3)$ instead of $L^1(\mathbb{R}^3)$, we have no guarantees that such line integrals even exist (converge). Putting this issue aside, note that in order to use the standard theory of linear least squares, we will also want the output of the operator L to lie

in a Hilbert space. Since Lf is a function over $\mathbb{R}^2 \times \mathbb{T}$, the natural space is

$$L^2(\mathbb{R}^2 \times \mathbb{T}) = \left\{ g : \mathbb{R}^2 \times \mathbb{T} \to \mathbb{R} \;\middle|\; \int_0^{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x, y, \theta)|^2 \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}\theta < \infty \right\},$$

where the inner product is defined as usual. Here, in order to use least squares, we need to go beyond our original assumption that Lf is a well-defined function of $\mathbb{R}^2 \times \mathbb{T}$ and make the further assumptions that $Lf \in L^2(\mathbb{R}^2 \times \mathbb{T})$ and moreover, that the operator L is a *bounded* linear operator, that is, L is a linear operator for which there exists some constant B such that $\|Lf\| \leq B\|f\|$ for all $f \in L^2(\mathbb{R}^3)$.

Making this assumption, we would like to find the f that solves the equation $Lf = g$ for a given g. However, the presence of noise in g makes it likely that this equation will not have a solution. As such, we instead attempt to solve

$$\operatorname*{arg\,min}_{f \in L^2(\mathbb{R}^3)} \|Lf - g\|_2^2. \tag{14}$$

That is, we want to find the f that minimizes the integral of the squares of the pointwise errors $(Lf)(x, y, \theta) - g(x, y, \theta)$. If we make another oversimplifying assumption—that the range of L is a *closed* subspace of $L^2(\mathbb{R}^2 \times \mathbb{T})$—classical Hilbert space theory implies that the solutions to (14) are equivalent to the solutions to the *normal equations* $L^*Lf = L^*g$.

The normal equations make use of the operator $L^* : L^2(\mathbb{R}^2 \times \mathbb{T}) \to L^2(\mathbb{R}^3)$, that is the adjoint of L, which is characterized by the fact that $\langle f, L^*g \rangle = \langle Lf, g \rangle$ for all $f \in L^2(\mathbb{R}^3)$ and $g \in L^2(\mathbb{R}^2 \times \mathbb{T})$. To see how to informally derive the formula

for L^* given in (6), consider

$$\begin{aligned}\langle f, L^* g\rangle &= \langle Lf, g\rangle \\ &= \int_0^{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [Lf(x,y,\theta)]^*\, g(x,y,\theta)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\theta \\ &= \int_0^{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x - z\cos\theta, y - z\sin\theta, z)\,\mathrm{d}z\right]^* g(x,y,\theta)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\theta.\end{aligned}$$

To proceed, we now distribute the complex conjugate and then interchange integrals (without justification) yielding

$$\langle f, L^* g\rangle = \int_0^{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [f(x - z\cos\theta, y - z\sin\theta, z)]^* g(x,y,\theta)\,\mathrm{d}z\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\theta.$$

For any fixed θ, making the substitution $(u, v, w) = (x - z\cos\theta, y - z\sin\theta, z)$ transforms this expression for $\langle f, L^* g\rangle$ into

$$\int_0^{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [f(u,v,w)]^* g(u + w\cos\theta, u + w\sin\theta, \theta)\mathrm{J}(u,v,w)\,\mathrm{d}w\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}\theta.$$

Here, $\mathrm{J}(u,v,w)$ is the Jacobian of this change of variables, namely the determinant of a 3×3 matrix, which happens to be upper triangular in this case:

$$\mathrm{J}(u,v,w) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix} = \begin{vmatrix} 1 & 0 & -\cos\theta \\ 0 & 1 & -\sin\theta \\ 0 & 0 & 1 \end{vmatrix} = (1)(1)(1) = 1.$$

Therefore, our expression for $\langle f, L^* g\rangle$ becomes

$$\langle f, L^* g\rangle = \int_0^{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [f(u,v,w)]^* g(u + w\cos\theta, u + w\sin\theta, \theta)\,\mathrm{d}w\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}\theta.$$

Once again, we interchange integrals (without justification) and find

$$\langle f, L^*g\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [f(u,v,w)]^* \left[\int_0^{2\pi} g(u + w\cos\theta, u + w\sin\theta, \theta)\,\mathrm{d}\theta\right] \mathrm{d}w\,\mathrm{d}u\,\mathrm{d}v.$$

That is, we necessarily have that, as given in (6), L^* has the form

$$(L^*g)(x,y,z) = \int_0^{2\pi} g(x + z\cos\theta, y + z\sin\theta, \theta)\,\mathrm{d}\theta.$$

In words, for each (x, y, z) triple, L^* integrates g over a helix. As noted above, L^* is commonly referred to as a back projection or as the shift-and-add method.

Returning to our goal of solving the normal equations $L^*Lf = L^*g$, we see that in order to reconstruct f, we first shift and add g and then attempt to invert the operator $L^*L : L^2(\mathbb{R}^3) \to L^2(\mathbb{R}^3)$. To do this, we now derive L^*L from (2) and (6):

$$\begin{aligned}
(L^*Lf)(x,y,z) &= \int_0^{2\pi} (Lf)(x + z\cos\theta, y + z\sin\theta, \theta)\,\mathrm{d}\theta \\
&= \int_0^{2\pi}\int_{-\infty}^{\infty} f(x + z\cos\theta - w\cos\theta, y + z\sin\theta - w\sin\theta, w)\,\mathrm{d}w\,d\theta \\
&= \int_0^{2\pi}\int_{-\infty}^{\infty} f(x + (z-w)\cos\theta, y + (z-w)\sin\theta, w)\,\mathrm{d}w\,\mathrm{d}\theta.
\end{aligned}$$

For any fixed θ, making the substitution $t = w - z$ gives

$$(L^*Lf)(x,y,z) = \int_0^{2\pi}\int_{-\infty}^{\infty} f(x - t\cos\theta, y - t\sin\theta, z + t)\,\mathrm{d}t\,\mathrm{d}\theta.$$

We now notice that L^*L is an example of a well-studied class of linear operators, namely *filters*. Filters arise from translations, which are operators that shift a function's input by a certain amount. In L^*L, the x, y, and z inputs are shifted by $t\cos\theta$,

$t \sin\theta$, and $-t$, respectively. It is well-known that such filters are *diagonalized* by the (continuous) Fourier transform, $E^* : L^2(\mathbb{R}^3) \to L^2(\mathbb{R}^3)$, defined as

$$(E^* f)(\alpha, \beta, \gamma) := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{e}^{-2\pi \mathrm{i}(\alpha x + \beta y + \gamma z)} f(x, y, z) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z.$$

The Fourier transform is a unitary operator, meaning that $E^* = E^{-1}$, where $E : L^2(\mathbb{R}^3) \to L^2(\mathbb{R}^3)$ is the Inverse Fourier Transform, defined as

$$(Ef)(x, y, z) := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{e}^{2\pi \mathrm{i}(\alpha x + \beta y + \gamma z)} f(\alpha, \beta, \gamma) \, \mathrm{d}\alpha \, \mathrm{d}\beta \, \mathrm{d}\gamma.$$

To see how the Fourier transform plays a role here, we now apply it to $L^* L f$:

$$\begin{aligned}
&(E^* L^* L f)(\alpha, \beta, \gamma) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{e}^{-2\pi \mathrm{i}(\alpha x + \beta y + \gamma z)} (L^* L f)(x, y, z) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{e}^{-2\pi \mathrm{i}(\alpha x + \beta y + \gamma z)} \int_{0}^{2\pi} \int_{-\infty}^{\infty} f(x - t\cos\theta, y - t\sin\theta, z + t) \, \mathrm{d}t \, \mathrm{d}\theta \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z.
\end{aligned}$$

Again interchanging integrals (without justification) this becomes

$$\int_{0}^{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{e}^{-2\pi \mathrm{i}(\alpha x + \beta y + \gamma z)} f(x - t\cos\theta, y - t\sin\theta, z + t) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \, \mathrm{d}t \, \mathrm{d}\theta.$$

At this point, for any fixed t and θ, we make the change of variables $(u, v, w) = (x - t\cos\theta, y - t\sin\theta, z + t)$. It turns out that the $\mathrm{J}(u, v, w)$ for this change of variables is 1, and so our expression for $(E^* L^* L f)(\alpha, \beta, \gamma)$ becomes

$$\int_{0}^{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{e}^{-2\pi \mathrm{i}(\alpha(u + t\cos\theta) + \beta(v + t\sin\theta) + \gamma(w - t))} f(u, v, w) \, \mathrm{d}u \, \mathrm{d}v \, \mathrm{d}w \, \mathrm{d}t \, \mathrm{d}\theta.$$

Once again, we switch the order of integration and our expression becomes

$$\begin{aligned}
&\int_0^{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-2\pi\mathrm{i}(\alpha t\cos\theta+\beta t\sin\theta-\gamma t)}\mathrm{d}t\,\mathrm{d}\theta \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-2\pi\mathrm{i}(\alpha u+\beta v+\gamma w)} f(u,v,w)\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w \\
&= \left(\int_0^{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-2\pi\mathrm{i}(\alpha t\cos\theta+\beta t\sin\theta+\gamma(-t))}\mathrm{d}t\,\mathrm{d}\theta\right)(E^*f)(\alpha,\beta,\gamma) \\
&= \lambda(\alpha,\beta,\gamma)(E^*f)(\alpha,\beta,\gamma),
\end{aligned}$$

where for any $(\alpha,\beta,\gamma)\in\mathbb{R}^3$, the scalar $\lambda(\alpha,\beta,\gamma)$ is defined as

$$\lambda(\alpha,\beta,\gamma) := \int_0^{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-2\pi\mathrm{i}(\alpha t\cos\theta+\beta t\sin\theta+\gamma(-t))}\mathrm{d}t\,\mathrm{d}\theta.$$

That is, we have $(E^*L^*Lf)(\alpha,\beta,\gamma) = \lambda(\alpha,\beta,\gamma)(E^*f)(\alpha,\beta,\gamma)$, meaning that in the Fourier domain, L^*L acts as a pointwise multiplication operator. In essence, for any $(\alpha,\beta,\gamma)\in\mathbb{R}^3$, we have that the function $\mathrm{e}_{\alpha,\beta,\gamma}(x,y,z) := \mathrm{e}^{2\pi\mathrm{i}(\alpha x+\beta y+\gamma z)}$ is an eigenfunction for L^*L with corresponding eigenvalue $\lambda(\alpha,\beta,\gamma)$. To understand this better, we now further investigate $\lambda(\alpha,\beta,\gamma)$. In particular, we write any $(\alpha,\beta)\in\mathbb{R}^2$ in polar form as

$$(\alpha,\beta) = \sqrt{\alpha^2+\beta^2}\left(\frac{\alpha}{\sqrt{\alpha^2+\beta^2}},\frac{\beta}{\sqrt{\alpha^2+\beta^2}}\right) = \sqrt{\alpha^2+\beta^2}(\cos\phi_{\alpha,\beta},\sin\phi_{\alpha,\beta}).$$

This allows us to make the following observation:

$$\begin{aligned}\lambda(\alpha,\beta,\gamma) &= \int_0^{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-2\pi\mathrm{i}(\alpha t\cos\theta+\beta t\sin\theta+\gamma(-t))}\mathrm{d}t\,\mathrm{d}\theta \\ &= \int_{-\infty}^{\infty} \mathrm{e}^{2\pi\mathrm{i}\gamma t}\int_0^{2\pi} \mathrm{e}^{-2\pi\mathrm{i}\sqrt{\alpha^2+\beta^2}t\left(\frac{\alpha}{\sqrt{\alpha^2+\beta^2}}\cos\theta+\frac{\beta}{\sqrt{\alpha^2+\beta^2}}\sin\theta\right)}\,\mathrm{d}\theta\ \mathrm{d}t \\ &= \int_{-\infty}^{\infty} \mathrm{e}^{2\pi\mathrm{i}\gamma t}\int_0^{2\pi} \mathrm{e}^{-2\pi\mathrm{i}\sqrt{\alpha^2+\beta^2}t\left(\cos\phi_{\alpha,\beta}\cos\theta+\sin\phi_{\alpha,\beta}\sin\theta\right)}\,\mathrm{d}\theta\ \mathrm{d}t.\end{aligned}$$

Using the angle sum formula for cosine, we find

$$\begin{aligned}\lambda(\alpha,\beta,\gamma) &= \int_{-\infty}^{\infty} \mathrm{e}^{2\pi\mathrm{i}\gamma t}\int_0^{2\pi} \mathrm{e}^{-2\pi\mathrm{i}\sqrt{\alpha^2+\beta^2}t\cos\left(\theta-\phi_{\alpha,\beta}\right)}\,\mathrm{d}\theta\ \mathrm{d}t \\ &= \int_{-\infty}^{\infty} \mathrm{e}^{2\pi\mathrm{i}\gamma t}\int_{-\pi}^{\pi} \mathrm{e}^{-2\pi\mathrm{i}\sqrt{\alpha^2+\beta^2}t\cos\theta}\,\mathrm{d}\theta\ \mathrm{d}t \\ &= 2\int_{-\infty}^{\infty} \mathrm{e}^{2\pi\mathrm{i}\gamma t}\int_0^{\pi} \mathrm{e}^{-2\pi\mathrm{i}\sqrt{\alpha^2+\beta^2}t\cos\theta}\,\mathrm{d}\theta\ \mathrm{d}t.\end{aligned}$$

Making the substitution $\phi = \theta - \frac{\pi}{2}$ transforms this expression into

$$\begin{aligned}\lambda(\alpha,\beta,\gamma) &= 2\int_{-\infty}^{\infty} \mathrm{e}^{2\pi\mathrm{i}\gamma t}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{e}^{-2\pi\mathrm{i}\sqrt{\alpha^2+\beta^2}t\cos(\phi+\frac{\pi}{2})}\,\mathrm{d}\phi\ \mathrm{d}t \\ &= 2\int_{-\infty}^{\infty} \mathrm{e}^{2\pi\mathrm{i}\gamma t}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{e}^{-2\pi\mathrm{i}\sqrt{\alpha^2+\beta^2}t(\cos\phi\cos\frac{\pi}{2}-\sin\phi\sin\frac{\pi}{2})}\,\mathrm{d}\phi\ \mathrm{d}t \\ &= 2\int_{-\infty}^{\infty} \mathrm{e}^{2\pi\mathrm{i}\gamma t}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{e}^{2\pi\mathrm{i}\sqrt{\alpha^2+\beta^2}t\sin\phi}\,\mathrm{d}\phi\ \mathrm{d}t.\end{aligned}$$

At this point, making the substitution $u = \sin\phi$, our expression becomes

$$\begin{aligned}\lambda(\alpha,\beta,\gamma) &= 2\int_{-\infty}^{\infty} \mathrm{e}^{2\pi \mathrm{i}\gamma t} \int_{-1}^{1} \mathrm{e}^{2\pi \mathrm{i}\sqrt{\alpha^2+\beta^2}tu} \frac{1}{\sqrt{1-u^2}}\,\mathrm{d}u\,\mathrm{d}t \\ &= 2\int_{-\infty}^{\infty} \mathrm{e}^{2\pi \mathrm{i}\gamma t} \int_{-\infty}^{\infty} \mathrm{e}^{2\pi \mathrm{i}\sqrt{\alpha^2+\beta^2}tu} \frac{1_{(-1,1)}(u)}{\sqrt{1-u^2}}\,\mathrm{d}u\,\mathrm{d}t.\end{aligned}$$

Letting $v = -u\sqrt{\alpha^2+\beta^2}$ transforms this expression into

$$\begin{aligned}\lambda(\alpha,\beta,\gamma) &= 2\int_{-\infty}^{\infty} \mathrm{e}^{2\pi \mathrm{i}\gamma t} \int_{-\infty}^{\infty} \mathrm{e}^{-2\pi \mathrm{i}tv} \frac{1_{(-1,1)}\left(\frac{-v}{\sqrt{\alpha^2+\beta^2}}\right)}{\sqrt{\alpha^2+\beta^2}\sqrt{1-\frac{v^2}{\alpha^2+\beta^2}}}\,\mathrm{d}v\,\mathrm{d}t \\ &= 2\int_{-\infty}^{\infty} \mathrm{e}^{2\pi \mathrm{i}\gamma t} \int_{-\infty}^{\infty} \mathrm{e}^{-2\pi \mathrm{i}tv} \frac{1_{(-1,1)}\left(\frac{v}{\sqrt{\alpha^2+\beta^2}}\right)}{\sqrt{\alpha^2+\beta^2-v^2}}\,\mathrm{d}v\,\mathrm{d}t.\end{aligned}$$

Recognizing this expression to be a composition of a one-dimensional Fourier transform with its inverse, our expression becomes

$$\begin{aligned}\lambda(\alpha,\beta,\gamma) &= 2\frac{1_{(-1,1)}\left(\frac{\gamma}{\sqrt{\alpha^2+\beta^2}}\right)}{\sqrt{\alpha^2+\beta^2-\gamma^2}} \\ &= \begin{cases} \frac{2}{\sqrt{\alpha^2+\beta^2-\gamma^2}}, & \gamma^2 < \alpha^2+\beta^2, \\ 0, & \gamma^2 \geq \alpha^2+\beta^2. \end{cases}\end{aligned}$$

Combining this fact with our earlier observation gives

$$(E^*L^*Lf)(\alpha,\beta,\gamma) = (E^*f)(\alpha,\beta,\gamma) \times \begin{cases} \frac{2}{\sqrt{\alpha^2+\beta^2-\gamma^2}}, & \gamma^2 < \alpha^2+\beta^2, \\ 0, & \gamma^2 \geq \alpha^2+\beta^2. \end{cases} \tag{15}$$

Note that $\lambda(\alpha,\beta,\gamma) = 0$ whenever $\gamma^2 \geq \alpha^2+\beta^2$. This means L^*L annihilates all the frequency information of f that lives in the cone $\{(\alpha,\beta,\gamma) : \gamma^2 \geq \alpha^2+\beta^2\}$. This is why this cone is called the "cone of missing information." This means that even

if all the assumptions we made in this chapter were true, in order to reconstruct f from g we would still be faced with solving a linear system $L^*Lf = L^*g$ where the null space of L^*L is gigantic. This means that in order to uniquely solve our system, we must make additional assumptions about f (for example, that f is sparse or its gradient is sparse), and then reconstruct according to these assumptions. In practice, such restrictions on f are usually enforced by regularizing our least squares problem and then performing the reconstruction according to some numerical optimization scheme. Since, at this point, all of our computations will necessarily become finite-dimensional, we are led to first investigate how the above derivations generalize to that setting. As we shall see, in our finite-dimensional CT model, we can rigorously address all of the technical issues that we oversimplified in this chapter. We now turn to a discrete, finite-dimensional model that will enable us to rigorously address all of the issues that were glossed over here.

III. A discrete mathematical model of chromotomography

Thus far, we have modeled our CT-based hyperspectral imaging system in terms of line integrals of functions over the continuum. However, as seen in the previous chapter, most of our derivations in these continuous models are not mathematically rigorous. In order to increase our understanding of our camera system and to make our arguments completely legitimate, we now discretize our hyperspectral image f, our operator L, and the surrounding theory.

Note that in the continuous model, in order to understand the null space of L^*L (and thus its noninvertibility), we eventually made use of Fourier transforms. Therefore, we want a function space over which a Fourier transform can be defined. Though many such spaces exist, only a select few of these are finite-dimensional, as needed in order to rigorously derive our CT theory using as little functional and harmonic analysis as possible. In particular, since our hyperspectral image $\mathbf{f}$ will, in this finite-dimensional model, become a discrete three-dimensional data cube, we use the three-dimensional Discrete Fourier Transform (DFT). This transform is defined over spaces of scalar-valued functions of $\mathbb{Z}^3 := \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$ that are periodic in each direction. To be precise, for any positive integer P, we consider the space

$$\ell(\mathbb{Z}_P^3) = \left\{ \mathbf{f} : \mathbb{Z}^3 \to \mathbb{C} \;\middle|\; \begin{aligned} \mathbf{f}[m,n,p] &= \mathbf{f}[m+P,n,p] \\ &= \mathbf{f}[m,n+P,p] \;\;,\;\; \forall m,n,p \in \mathbb{Z} \\ &= \mathbf{f}[m,n,p+P] \end{aligned} \right\}.$$

It is well-known that $\ell(\mathbb{Z}_P^3)$ space is a Hilbert space under the inner product

$$\langle \mathbf{f}_1, \mathbf{f}_2 \rangle := \sum_{m \in \mathbb{Z}_P} \sum_{n \in \mathbb{Z}_P} \sum_{p \in \mathbb{Z}_P} (\mathbf{f}_1[m,n,p])^* \mathbf{f}_2[m,n,p].$$

Here and throughout, summing over all "$m \in \mathbb{Z}_P$" means to sum over a set of coset representatives of the subgroup $P\mathbb{Z}$ of the integers $\mathbb{Z}$. For example, one choice of coset representatives is to take $m = 0, ..., P-1$. Note this space has dimension P^3,

with the only independent quantities of $\mathbf{f}$ being the values $\mathbf{f}[m,n,p]$ for $m,n,p = 0,1,\ldots,P-1$. Recall that in our continuous model, light at frequency $z=0$ passes directly through the prism. Also, when the prism orientation angle θ is 0, light at frequencies $z=1$ and $z=-1$ is shifted one unit to the right and left, respectively. In order to keep this same intuition in the discrete setting, it is helpful to regard 0 as the center of $\mathbb{Z}_P$. That is, when P is even, we take the coordinates of $\mathbf{f}$ to range from $-\frac{P}{2}$ to $\frac{P}{2}-1$. Meanwhile when P is odd, we take them from $-\frac{P-1}{2}$ to $\frac{P-1}{2}$.

As seen in (2), the continuous X-ray transform of $f \in L^2(\mathbb{R}^3)$ at a given triple $(x,y,\theta) \in \mathbb{R}^2 \times \mathbb{T}$ is obtained by integrating $f(x - z\cos\theta, y - z\sin\theta, z)$ over every frequency $z \in \mathbb{R}$. In the discrete setting, the input for our operator $\mathbf{L}$ will be a triple $[m,n,q]$ where m and n are the horizontal and vertical pixel indices, respectively, and q represents a discrete angle parameter. Specifically, given some positive integer Q, let $\psi_1, \psi_2 \in \ell(\mathbb{Z}_Q, \mathbb{Z})$ be two Q-periodic integer-valued functions defined over $\mathbb{Z}$. These functions will play the role of discrete cosines and sines: below, when we discretize our continuous X-ray transform (2), we will replace the terms $z\cos\theta$ and $z\sin\theta$ with $p\psi_1[q]$ and $p\psi_2[q]$, respectively.

For example, if our camera system used $Q=4$ angles, we may use the four evenly-spaced pairs of integers, namely

$$\begin{aligned} (\psi_1[0], \psi_2[0]) &= (1,0), \\ (\psi_1[1], \psi_2[1]) &= (0,1), \\ (\psi_1[2], \psi_2[2]) &= (-1,0), \\ (\psi_1[3], \psi_2[3]) &= (0,-1). \end{aligned} \tag{16}$$

If we wanted to use $Q=8$ angles, we could use the "knight's moves" set:

$$\{(2,1),(1,2),(-1,2),(-2,1),(-2,-1),(-1,-2),(1,-2),(2,-1)\}. \tag{17}$$

These 8 points are depicted in Figure 1. We can now form a discrete operator $\mathbf{L}$

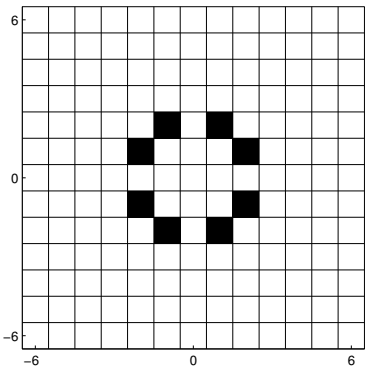

Figure 1: Here we portray the 8 symmetric discrete "angles" known as the "knight's moves" set, as given in (17). For this figure, we have chosen $P = 13$.

which generalizes (2):

$$(\mathbf{Lf})[m,n,q] = \sum_{p\in\mathbb{Z}_P} \mathbf{f}[m - p\psi_1[q], n - p\psi_2[q], p]. \tag{18}$$

That is, instead of integrating over all frequencies z, we now sum over a finite number of frequencies $p \in \mathbb{Z}_P$. Note that since $\mathbf{f} \in \ell(\mathbb{Z}_P^3)$ and $\psi_1, \psi_2 \in \ell(\mathbb{Z}_Q, \mathbb{Z})$, then $\mathbf{Lf}$ is (P, P, Q)-periodic. Indeed, the P-periodicity of $\mathbf{f}$ gives

$$\begin{aligned}(\mathbf{Lf})[m+P,n,q] &= \sum_{p\in\mathbb{Z}_P} \mathbf{f}[m + P - p\psi_1[q], n - p\psi_2[q], p] \\ &= \sum_{p\in\mathbb{Z}_P} \mathbf{f}[m - p\psi_1[q], n - p\psi_2[q], p] \\ &= (\mathbf{Lf})[m,n,q].\end{aligned}$$

The fact that $(\mathbf{Lf})[m, n+P, q] = (\mathbf{Lf})[m, n, q]$ follows similarly. Finally, the Q-periodicity of ψ_1 and ψ_2 gives

$$\begin{aligned}(\mathbf{Lf})[m,n,q+Q] &= \sum_{p\in\mathbb{Z}_P} \mathbf{f}[m - p\psi_1[q+Q], n - p\psi_2[q+Q], p] \\ &= \sum_{p\in\mathbb{Z}_P} \mathbf{f}[m - p\psi_1[q], n - p\psi_2[q], p] \\ &= (\mathbf{Lf})[m,n,q].\end{aligned}$$

That is, $\mathbf{Lf}$ lies in the Hilbert space

$$\ell(\mathbb{Z}_P^2 \times \mathbb{Z}_Q) = \left\{ \mathbf{g} : \mathbb{Z}^3 \to \mathbb{C} \;\middle|\; \begin{array}{rcl} \mathbf{g}[m,n,q] &=& \mathbf{g}[m+P,n,q] \\ &=& \mathbf{g}[m,n+P,q] \;\;,\;\; \forall m,n,q \in \mathbb{Z} \\ &=& \mathbf{g}[m,n,q+Q] \end{array} \right\}.$$

One can immediately see that $\mathbf{L} : \ell(\mathbb{Z}_P^3) \to \ell(\mathbb{Z}_P^2 \times \mathbb{Z}_Q)$ is linear. Also, since $\ell(\mathbb{Z}_P^3)$ is finite-dimensional, this linear operator is automatically bounded. Moreover, since $\mathbf{L}$ is bounded its adjoint exists. As such, the technicalities we encountered in Chapter II concerning whether or not $\mathbf{L}$ and $\mathbf{L}^*$ are well-defined do not arise here.

Having that $\mathbf{L}^*$ exists, we now parallel the derivation in the continuous setting in order to find an explicit formula for it. Throughout the following calculations, summations over the variables m, n, p range over $\mathbb{Z}_P$, while those over the variable q range over $\mathbb{Z}_Q$. For any $\mathbf{f} \in \ell(\mathbb{Z}_P^3)$ and any $\mathbf{g} \in \ell(\mathbb{Z}_P^2 \times \mathbb{Z}_Q)$:

$$\begin{aligned} \langle \mathbf{f}, \mathbf{L}^*\mathbf{g} \rangle &= \langle \mathbf{Lf}, \mathbf{g} \rangle \\ &= \sum_m \sum_n \sum_q ((\mathbf{Lf})[m,n,q])^* \mathbf{g}[m,n,q] \\ &= \sum_m \sum_n \sum_q \sum_p (\mathbf{f}[m - p\psi_1[q], n - p\psi_2[q], p])^* \mathbf{g}[m,n,q]. \end{aligned}$$

At this point in Chapter II, we interchanged integrals without justification. Here however, our sums are finite, and so we can interchange them freely:

$$\langle \mathbf{f}, \mathbf{L}^*\mathbf{g} \rangle = \sum_q \sum_p \sum_m \sum_n (\mathbf{f}[m - p\psi_1[q], n - p\psi_2[q], p])^* \mathbf{g}[m,n,q].$$

For any p and q, we now substitute (m', n') for $(m - p\psi_1[q], n - p\psi_1[q])$:

$$\begin{aligned}\langle \mathbf{f}, \mathbf{L}^*\mathbf{g} \rangle &= \sum_q \sum_p \sum_{m'} \sum_{n'} (\mathbf{f}[m', n', p])^* \mathbf{g}[m' + p\psi_1[q], n' + p\psi_2[q], q] \\ &= \sum_{m'} \sum_{n'} \sum_p (\mathbf{f}[m', n', p])^* \sum_q \mathbf{g}[m' + p\psi_1[q], n' + p\psi_2[q], q].\end{aligned}$$

That is, the adjoint of $\mathbf{L}$ is

$$(\mathbf{L}^*\mathbf{g})[m, n, p] = \sum_{q \in \mathbb{Z}_Q} \mathbf{g}[m + p\psi_1[q], n + p\psi_2[q], q]. \tag{19}$$

Note the similarities between (6) and (19); the $\mathbf{L}^*$ given here adds up the values of $\mathbf{g}$ over a discrete "helix" centered at $[m, n]$ with radius p.

At this point, recall that we want $\mathbf{L}^*$ in order to help us solve the least squares problem $\arg\min_{\mathbf{f}} \|\mathbf{L}\mathbf{f} - \mathbf{g}\|_2^2$. Since $\mathbf{L}$ is a bounded linear operator between Hilbert spaces, this problem is equivalent to solving the normal equations $\mathbf{L}^*\mathbf{L}\mathbf{f} = \mathbf{L}^*\mathbf{g}$. Moreover, these equations will have a solution provided the range of $\mathbf{L}$ is *closed*. In the continuous setting, this was yet another assumption we made in order to proceed. Here however, we need not make such an assumption since it is automatically true: the range of $\mathbf{L}$ is a subspace of the *finite*-dimensional space $\ell(\mathbb{Z}_P^2 \times \mathbb{Z}_Q)$ and so is necessarily closed. Thus, we can indeed minimize $\|\mathbf{L}\mathbf{f} - \mathbf{g}\|_2^2$ by solving $\mathbf{L}^*\mathbf{L}\mathbf{f} = \mathbf{L}^*\mathbf{g}$.

That is, given camera data $\mathbf{g} \in \ell(\mathbb{Z}_P^2 \times \mathbb{Z}_Q)$, our first step is to compute $\mathbf{L}^*\mathbf{g}$ according to (19). We then attempt to solve the normal equations. To facilitate this, we now derive an explicit formula for $\mathbf{L}^*\mathbf{L}$ from (18) and (19):

$$\begin{aligned}(\mathbf{L}^*\mathbf{L}\mathbf{f})[m, n, p] &= \sum_q (\mathbf{L}\mathbf{f})[m + p\psi_1[q], n + p\psi_2[q], q] \\ &= \sum_q \sum_{p'} \mathbf{f}[m + p\psi_1[q] - p'\psi_1[q], n + p\psi_2[q] - p'\psi_2[q], p'].\end{aligned}$$

For any fixed p, letting $r = p' - p$ gives

$$(\mathbf{L}^*\mathbf{L}\mathbf{f})[m,n,p] = \sum_{q\in\mathbb{Z}_Q}\sum_{r\in\mathbb{Z}_P} \mathbf{f}[m - r\psi_1[q], n - r\psi_2[q], p + r]. \tag{20}$$

In Chapter II, we informally discussed how our continuous operator $\mathbf{L}^*\mathbf{L}$ was a filter. Now that we are in the discrete setting, we can make this notion rigorous by using the theory of *convolutions*. Convolution is a type of vector-vector product: it takes any two functions and produces a third. To be precise, we define the convolution of $\mathbf{f}_1, \mathbf{f}_2 \in \ell(\mathbb{Z}_P^3)$ as

$$(\mathbf{f}_1 * \mathbf{f}_2)[m,n,p] = \sum_{m'\in\mathbb{Z}_P}\sum_{n'\in\mathbb{Z}_P}\sum_{p'\in\mathbb{Z}_P} \mathbf{f}_1[m',n',p']\mathbf{f}_2[m-m', n-n', p-p']. \tag{21}$$

This convolutional product is well-known, having many nice properties. In particular, under standard addition and this notion of multiplication, it turns out that $\ell(\mathbb{Z}_P^3)$ is a commutative ring with identity $\boldsymbol{\delta}_{(0,0,0)}$. Here, for any $m_0, n_0, p_0 \in \mathbb{Z}$, the *Dirac-$\boldsymbol{\delta}$* function at (m_0, n_0, p_0) is $\boldsymbol{\delta}_{(m_0,n_0,p_0)} \in \ell(\mathbb{Z}_P^3)$,

$$\boldsymbol{\delta}_{(m_0,n_0,p_0)}[m,n,p] := \begin{cases} 1, & m = m_0,\ n = n_0,\ p = p_0, \quad \text{mod } P, \\ 0, & \text{else.} \end{cases}$$

One way to think about the convolution of $\mathbf{f}_1$ and $\mathbf{f}_2$ is that it is a "rolling average" of the values of $\mathbf{f}_1$ with coefficients from $\mathbf{f}_2$. As we now show, this is

precisely what our operator $\mathbf{L}^*\mathbf{L}$ does; we rewrite (20) as

$$\begin{aligned}(\mathbf{L}^*\mathbf{L}\mathbf{f})[m,n,p] &= \sum_q \sum_r \mathbf{f}[m - r\psi_1[q], n - r\psi_2[q], p - (-r)] \\ &= \sum_q \sum_r (\boldsymbol{\delta}_{(r\psi_1[q], r\psi_2[q], -r)} * \mathbf{f})[m,n,p] \\ &= \left(\sum_q \sum_r \boldsymbol{\delta}_{r(\psi_1[q], \psi_2[q], -1)} \right) * \mathbf{f}[m,n,p] \\ &= \left(\sum_q \mathbf{h}_q \right) * \mathbf{f}[m,n,p]\end{aligned}$$

where, for any $q \in \mathbb{Z}$, $\mathbf{h}_q \in \ell(\mathbb{Z}_P^3)$ is a "characteristic function" of the discrete "line" that consists of all integer multiples of $(\psi_1[q], \psi_2[q], -1)$:

$$\mathbf{h}_q := \sum_{r \in \mathbb{Z}_P} \boldsymbol{\delta}_{r(\psi_1[q], \psi_2[q], -1)}. \tag{22}$$

We can now express $\mathbf{L}^*\mathbf{L}$ as

$$\mathbf{L}^*\mathbf{L}\mathbf{f} = \mathbf{h} * \mathbf{f}, \tag{23}$$

where $\mathbf{h} \in \ell(\mathbb{Z}_P^3)$ represents the "surface" of a discrete "cone":

$$\mathbf{h}[m,n,p] := \sum_{q \in \mathbb{Z}_Q} \mathbf{h}_q[m,n,p] = \sum_{q \in \mathbb{Z}_Q} \sum_{r \in \mathbb{Z}_P} \boldsymbol{\delta}_{r(\psi_1[q], \psi_2[q], -1)}[m,n,p]. \tag{24}$$

That is, at any $(m,n,p) \in \mathbb{Z}^3$, $\mathbf{L}^*\mathbf{L}\mathbf{f}$ represents an "average" of those values of $\mathbf{f}$ that lie on any one of the Q discrete "lines" of all integer multiples of $(\psi_1[q], \psi_2[q], -1)$. For example, when $Q = 8$ and ψ_1, ψ_2 correspond to the knight's moves of (17), $(\mathbf{L}^*\mathbf{L}\mathbf{f})[m,n,p]$ adds up the values of $\mathbf{f}$ over the surface of a "cone" consisting of 8 "lines" radiating outward from (m,n,p).

Having that $\mathbf{L}^*\mathbf{L}$ is a filter, namely, the convolutional operator $\mathbf{L}^*\mathbf{L}\mathbf{f} = \mathbf{h} * \mathbf{f}$, we can now better understand it using Fourier transforms. In particular, since our model is now discrete, we use the Discrete Fourier Transform (DFT), namely the

operator $\mathbf{E}^* : \ell(\mathbb{Z}_P^3) \to \ell(\mathbb{Z}_P^3)$, defined as

$$(\mathbf{E}^*\mathbf{f})[\alpha,\beta,\gamma] := \sum_{m\in\mathbb{Z}_P}\sum_{n\in\mathbb{Z}_P}\sum_{p\in\mathbb{Z}_P} \mathrm{e}^{-\frac{2\pi\mathrm{i}}{P}(\alpha m+\beta n+\gamma p)}\mathbf{f}[m,n,p].$$

It is well-known that the DFT is a scalar multiple of a unitary operator, meaning its inverse is a scalar multiple of its adjoint. In particular, $(\mathbf{E}^*)^{-1} = \frac{1}{P^3}\mathbf{E}$, where $\mathbf{E} : \ell(\mathbb{Z}_P^3) \to \ell(\mathbb{Z}_P^3)$,

$$(\mathbf{E}\mathbf{f})[m,n,p] = \sum_{\alpha\in\mathbb{Z}_P}\sum_{\beta\in\mathbb{Z}_P}\sum_{\gamma\in\mathbb{Z}_P} \mathrm{e}^{\frac{2\pi\mathrm{i}}{P}(\alpha m+\beta n+\gamma p)}\mathbf{f}[\alpha,\beta,\gamma].$$

It is also well-known that the DFT of a discrete convolution (21) is a scalar multiple of the pointwise product of their individual DFTs:

$$(\mathbf{E}^*(\mathbf{f}_1 * \mathbf{f}_2))[\alpha,\beta,\gamma] = (\mathbf{E}^*\mathbf{f}_1)[\alpha,\beta,\gamma](\mathbf{E}^*\mathbf{f}_2)[\alpha,\beta,\gamma].$$

Applying this fact to $\mathbf{L}^*\mathbf{L}\mathbf{f} = \mathbf{h} * \mathbf{f}$ gives

$$(\mathbf{E}^*\mathbf{L}^*\mathbf{L}\mathbf{f})[\alpha,\beta,\gamma] = (\mathbf{E}^*\mathbf{h})[\alpha,\beta,\gamma]\,(\mathbf{E}^*\mathbf{f})\,[\alpha,\beta,\gamma].$$

Letting $\boldsymbol{\lambda} = (\mathbf{E}^*\mathbf{h})$ allows us to express the DFT of $\mathbf{L}^*\mathbf{L}\mathbf{f}$ as

$$(\mathbf{E}^*\mathbf{L}^*\mathbf{L}\mathbf{f})[\alpha,\beta,\gamma] = \boldsymbol{\lambda}[\alpha,\beta,\gamma](\mathbf{E}^*\mathbf{f})[\alpha,\beta,\gamma].$$

That is, just like in the continuous setting, $\mathbf{L}^*\mathbf{L}$ acts as a pointwise multiplication operator in the Fourier domain. Here, for any $(\alpha,\beta,\gamma) \in \mathbb{Z}^3$, $\mathbf{e}_{\alpha,\beta,\gamma} \in \ell(\mathbb{Z}_P^3)$, $\mathbf{e}_{\alpha,\beta,\gamma}[m,n,p] = \mathrm{e}^{\frac{2\pi\mathrm{i}}{P}(\alpha m+\beta n+\gamma p)}$ is an eigenfunction for $\mathbf{L}^*\mathbf{L}$ with corresponding eigenvalue $\boldsymbol{\lambda}[\alpha,\beta,\gamma]$. This is equivalent to us having $\mathbf{L}^*\mathbf{L} = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^*$ where $\boldsymbol{\Lambda} : \ell(\mathbb{Z}_P^3) \to \ell(\mathbb{Z}_P^3)$ is pointwise multiplication by $\boldsymbol{\lambda}$. To better understand the ramifications of

this fact, we now obtain a simplified expression for $\boldsymbol{\lambda}[\alpha, \beta, \gamma]$:

$$\begin{aligned}\boldsymbol{\lambda}[\alpha, \beta, \gamma] &= (\mathbf{E}^* \mathbf{h})[\alpha, \beta, \gamma] \\ &= (\mathbf{E}^* \sum_q \mathbf{h}_q)[\alpha, \beta, \gamma] \\ &= \sum_q (\mathbf{E}^* \mathbf{h}_q)[\alpha, \beta, \gamma].\end{aligned}$$

From here we will focus on calculating $\boldsymbol{\lambda}_q[\alpha, \beta, \gamma] = (\mathbf{E}^* \mathbf{h}_q)[\alpha, \beta, \gamma]$:

$$\begin{aligned}\boldsymbol{\lambda}_q[\alpha, \beta, \gamma] &= \left(\mathbf{E}^* \sum_r \boldsymbol{\delta}_{r(\psi_1[q], \psi_2[q], -1)} \right) [\alpha, \beta, \gamma] \\ &= \sum_m \sum_n \sum_p \mathrm{e}^{-\frac{2\pi i}{P}(\alpha m + \beta n + \gamma p)} \sum_r \boldsymbol{\delta}_{r(\psi_1[q], \psi_2[q], -1)}[m, n, p].\end{aligned}$$

To proceed, we interchange sums:

$$\begin{aligned}\boldsymbol{\lambda}_q[\alpha, \beta, \gamma] &= \sum_r \sum_m \sum_n \sum_p \mathrm{e}^{-\frac{2\pi i}{P}(\alpha m + \beta n + \gamma p)} \boldsymbol{\delta}_{r(\psi_1[q], \psi_2[q], -1)}[m, n, p] \\ &= \sum_r \mathrm{e}^{-\frac{2\pi i}{P}(\alpha r \psi_1[q] + \beta r \psi_2[q] + \gamma(-r))} \\ &= \sum_r \left[\mathrm{e}^{-\frac{2\pi i}{P}(\alpha \psi_1[q] + \beta \psi_2[q] + \gamma(-1))} \right]^r.\end{aligned}$$

Using the Geometric Sum Formula, we end up with

$$\begin{aligned}\boldsymbol{\lambda}_q[\alpha, \beta, \gamma] &= \begin{cases} P, & \alpha \psi_1[q] + \beta \psi_2[q] - \gamma = 0 \mod P, \\ 0, & \text{else}, \end{cases} \\ &= P \begin{cases} 1, & \gamma = \alpha \psi_1[q] + \beta \psi_2[q] \mod P, \\ 0, & \text{else}. \end{cases}\end{aligned} \tag{25}$$

Note that the multiplier of P in (25) is the characteristic function of the plane in $\mathbb{Z}_P^3$ which is "perpendicular" to $(\psi_1[q], \psi_2[q], -1)$. We now summarize the main results of this chapter in the following theorem:

Theorem 1. *Let* $\mathbf{L} : \ell(\mathbb{Z}_P^3) \to \ell(\mathbb{Z}_P^2 \times \mathbb{Z}_Q)$ *be the discrete X-ray transform:*

$$(\mathbf{Lf})[m, n, q] := \sum_{p \in \mathbb{Z}_P} \mathbf{f}[m - p\psi_1[q], n - p\psi_2[q], p].$$

The adjoint of $\mathbf{L}$ *is* $\mathbf{L}^* : \ell(\mathbb{Z}_P^2 \times \mathbb{Z}_Q) \to \ell(\mathbb{Z}_P^3)$,

$$(\mathbf{L}^*\mathbf{g})[m, n, p] = \sum_{q \in \mathbb{Z}_Q} \mathbf{g}[m + p\psi_1[q], n + p\psi_2[q], q].$$

Moreover, $\mathbf{L}^*\mathbf{L} : \ell(\mathbb{Z}_P^3) \to \ell(\mathbb{Z}_P^3 \times \mathbb{Z}_Q)$*, is a filter, with* $\mathbf{L}^*\mathbf{Lf} = \mathbf{h} * \mathbf{f}$*, where*

$$\mathbf{h}[m, n, p] = \sum_{q \in \mathbb{Z}_Q} \sum_{r \in \mathbb{Z}_P} \boldsymbol{\delta}_{r(\psi_1[q], \psi_2[q], -1)}[m, n, p].$$

As such, the DFT of $\mathbf{L}^*\mathbf{Lf}$ *is* $(\mathbf{E}^*\mathbf{L}^*\mathbf{Lf})[\alpha, \beta, \gamma] = \boldsymbol{\lambda}[\alpha, \beta, \gamma](\mathbf{E}^*\mathbf{f})[\alpha, \beta, \gamma]$*, where*

$$\begin{aligned}\boldsymbol{\lambda}[\alpha, \beta, \gamma] &= \sum_{q \in \mathbb{Z}_Q} (\mathbf{E}^*\mathbf{h}_q)[\alpha, \beta, \gamma] \\ &= P \sum_{q \in \mathbb{Z}_Q} \begin{cases} 1, & \gamma = \alpha\psi_1[q] + \beta\psi_2[q] \ \textit{mod } P, \\ 0, & \textit{else.} \end{cases}\end{aligned}$$

As in the continuous setting, we see that $\mathbf{L}^*\mathbf{L}$ is not invertible. Indeed, the null space of $\mathbf{L}^*\mathbf{L}$ (which equals the null space of $\mathbf{L}$) consists of all $\mathbf{f} \in \ell(\mathbb{Z}_P^3)$ whose DFT is completely supported on the set-complement of every plane of the form $\{(\alpha, \beta, \gamma) \in \mathbb{Z}_P^3 : \gamma = \alpha\psi_1[q] + \beta\psi_2[q] \mod P\}$. Part of this null space corresponds to a discrete "cone of missing information." For example, when $Q = 4$ and ψ_1, ψ_2 are given by (16), the null space of $\mathbf{L}$ is the set of all $\mathbf{f}$ whose DFTs are identically zero on the four planes in $\mathbb{Z}_P^3$ consisting of the points of the form (α, β, α), (α, β, β),

$(\alpha, \beta, -\alpha)$, and $(\alpha, \beta, -\beta)$. The four planes form a rough "cone." Any $\mathbf{f}$ whose DFT is completely supported on the "interior" of this cone, namely the set $\{(\alpha, \beta, \gamma) : |\gamma| \leq \min\{|\alpha|, |\beta|\}\}$, lies in the null space of $\mathbf{L}$. All such $\mathbf{f}$'s are completely annihilated by $\mathbf{L}$, meaning any solution to $\mathbf{L}^*\mathbf{L}\mathbf{f} = \mathbf{L}^*\mathbf{g}$ is only unique up to the addition of such functions. Because of this non-uniqueness we, in Chapter V, will regularize our least squares problem with penalty terms that will promote those solutions to the normal equations which are either sparse or have sparse gradients.

For now however, we turn our attention to a more in-depth comparison of the above discrete model and the continuous model from Chapter II. We find that despite being fully rigorous, the discrete model is not entirely faithful to the motivating real-world physics, and as such, take steps to improve it.

IV. Approximating continuous chromotomography with a generalized discrete model

In Chapter II we introduced the linear operator (2) that provides a continuous-variable model of the CT camera system. In this model, the operator L integrates a hyperspectral image f over every line of the form $\{(x - z\cos\theta, y - z\sin\theta, z) : z \in \mathbb{R}\}$. The xy-projection of an example of such a line is depicted in Figure 2(a). Here θ is chosen so that this line has slope $\frac{1}{2}$, that is, so that $\tan\theta = \frac{1}{2}$. Every point on the line seen in Figure 2(a) contributes to our measurement $(Lf)(x, y, \theta)$. In Chapter III, we created a discrete version (18) of the X-ray transform. This operator behaves similarly to the continuous version in that it sums the values of a data cube $\mathbf{f}$ over discrete "lines," namely over all points of the form $\{(m - p\psi_1[q], n - p\psi_2[q], p) : p \in \mathbb{Z}_P\}$. Figure 2(b) depicts these points for a discrete "line" that has the same slope as the continuous line of Figure 2(a), namely one in which $\psi_1[q] = 2$, $\psi_2[q] = 1$.

Note there are two main differences between the continuous version seen in Figure 2(a) and the discrete version depicted in Figure 2(b). The first difference is that due to the periodicity in the discrete model, the discrete line "wraps around" the edges of our grid. We do not address this issue here. The second difference is that the discrete operator only takes into account the *integer* multiples of $(\psi_1[q], \psi_2[q], -1)$, and therefore "skips over" the pixels that lie between $p(\psi_1[q], \psi_2[q], -1)$ and $(p + 1)(\psi_1[q], \psi_2[q], -1)$. This chapter attempts to improve our discrete model by taking these "skipped" pixels into account.

In particular, rather than attempt to fully reconcile the discrete operator $\mathbf{L}$ with the continuous one, L, we instead try to relate $\mathbf{L}$ to a periodic analog of L. To be precise, let $\mathbb{T}_P := \mathbb{R}/(P\mathbb{Z})$ be the real numbers modulo P. Recall the operator L is applied to functions $f : \mathbb{R}^3 \to \mathbb{R}$. We now generalize it to another operator L_P that is applied to the P-periodic functions $f : \mathbb{T}_P^3 \to \mathbb{R}$. To do this, recall that L integrates f over lines of the form $\{(x - z\cos\theta, y - z\sin\theta, z) : z \in \mathbb{R}\}$. In order for the image of these lines in $\mathbb{T}_P^3$ to have finite length, it turns out that we must only

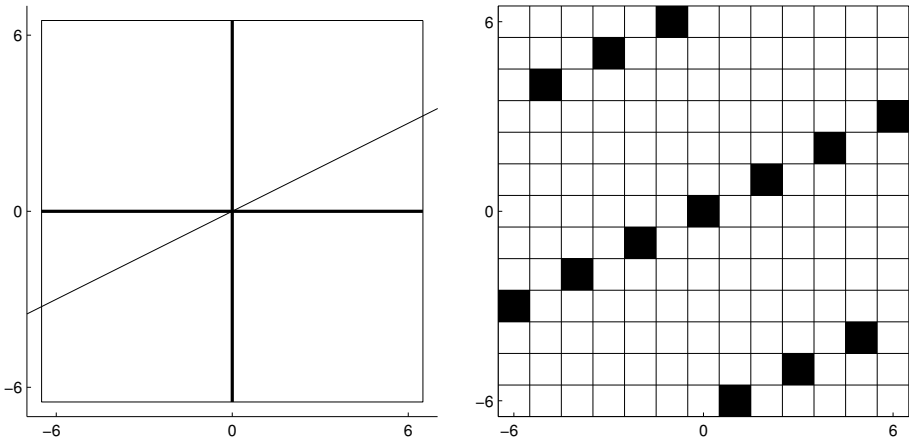

Figure 2: **(a)** In Chapter II, we investigate the continuous X-ray transform (2) which, of all the models presented here, provides the most accurate representation of the underlying real world physics of the CT camera system. This operator L integrates over lines of the form $(x - z\cos\theta, y - \sin\theta, z)$. In this figure we depict the xy-projection of one of these three-dimensional lines, namely the one in which $\tan\theta = \frac{1}{2}$ and $(x, y) = (0, 0)$. In Chapter II, we discuss how L^*L is a filter and can thus be diagonalized by the continuous Fourier transform. Examining L^*L in the Fourier domain allows us to identify the "cone of missing information," that is, the frequency information of f annihilated by L^*L. Unfortunately, it is difficult to make the theory from that chapter rigorous, motivating us to turn to the discrete theory in Chapter III. **(b)** In Chapter III we create a discrete version of L, that is, we create the operator $\mathbf{L}$ that sums over discrete "lines" of the form $(m - p\psi_1[q], n - p\psi_2[q], p)$. In this figure we depict a discrete version of the continuous line depicted in (a), namely one in which $(\psi_1[q], \psi_2[q]) = (2, 1)$. In Chapter III, we show that $\mathbf{L}^*\mathbf{L}$ convolves $\mathbf{f}$ with a sum of characteristic functions of lines of this form and can be understood by applying the DFT to it. In a similar manner to Chapter II, investigating $\mathbf{L}^*\mathbf{L}$ in the Fourier domain, we see that $\mathbf{L}^*\mathbf{L}$ eliminates the frequency components of $\mathbf{f}$ whose DFT is completely supported on the complement of every plane of the form $\gamma = \alpha\psi_1[q] + \beta\psi_2[q] \mod P$. This roughly corresponds to a discrete "cone of missing information." In this chapter, we focus on generalizing the discrete theory of Chapter III so as to make it better approximate the continuous theory of Chapter II. To do this, we modify the discrete "line" in (b) to make it look more like the continuous line (a).

use angles θ for which $\tan\theta$ is rational. That is, we take L_P to integrate over lines of the form $\{(x - z\psi_1[q], y - z\psi_2[q], z) : z \in \mathbb{T}_P\}$ where, for any angle parameter q, we require $\psi_1[q], \psi_2[q] \in \mathbb{Z}$. That is, for any $f \in L^2(\mathbb{T}_P^3)$ we let

$$(L_P f)(x, y, q) := \int_{\mathbb{T}_P} f(x - z\psi_1[q], y - z\psi_2[q], z)\,\mathrm{d}z. \tag{26}$$

For example, for $P = 13$ and $(\psi_1[q], \psi_2[q]) = (2, 1)$, the line that L_P integrates over is depicted in Figure 3(a). To see how a continuous integral over such a line can be related to a discrete sum over the "line" in Figure 2(b), it is helpful to overlay Figure 3(a) with a grid of pixels; see Figure 3(b). In particular, we claim that if f is

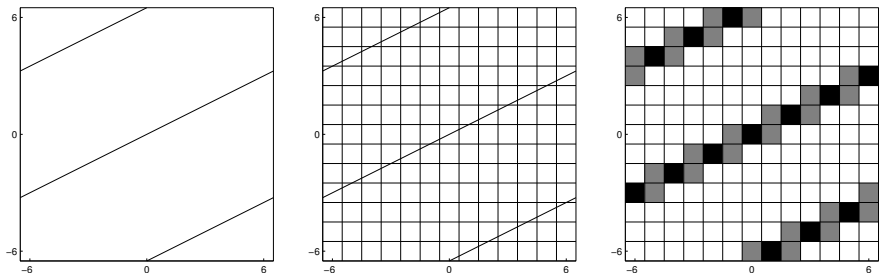

Figure 3: **(a)** To reconcile the operators L and $\mathbf{L}$ alluded to in Figure 2, we introduce a periodic version of the continuous X-ray transform. This continuous, periodic operator L_P, formally defined in (30), integrates over lines on a torus, which forces the "wrap around" effect, depicted here. While this model is not as physically realistic as L—space and color are not cyclic—it seems more realistic than the discrete periodic operator $\mathbf{L}$. However, this model is not ideal because, being analog rather than digital, it cannot exploit modern computational hardware. In order to make this continuous-periodic model more conducive for such computations, we, in Chapter IV, take steps to bridge the gap between it and our non-realistic but computational-friendly discrete-periodic operator $\mathbf{L}$. **(b)** Intuitively, if a continuous, P-periodic function f were sampled finely enough, we should be able to discretize its X-ray transform $L_P f$ in terms of finite sums. Graphically, we can think of this process as overlaying our continuous torus with an integer grid, as depicted here. Each square represents a region over which we assume our continuous function f assumes a constant value. Identifying this continuous function with its integer samples $\mathbf{f}$ we, in Theorem 2, show that the continuous-periodic X-ray transform of f indeed equals a weighted version of the discrete-periodic X-ray transform of $\mathbf{f}$. **(c)** In order to relate $L_P f$ to the discrete theory of Chapter III, it turns out it is necessary to generalize our discrete periodic X-ray transform $\mathbf{L}$ to another operator $\mathbf{K}$ that contains additional weight parameters $\mathbf{w}[a, b]$. As explained in Theorem 2, these weights are related to the proportion of our continuous line that lives in any given square. In essence, we form a discretized version of (a) by overlaying it with a grid (b). This results in a weight function (c). That is, when compared to a sum over the points depicted in Figure 2(b), summing over the weights in (c) does a better job of approximating the integral over the line depicted in (a). After Theorem 2, the remainder of this chapter is devoted to showing that $\mathbf{K}$ shares many of the nice computational properties of $\mathbf{L}$. That is, the point of this chapter is to show how, under slight modification, the discrete perspective of Chapter III can indeed be used to form a computationally efficient, mathematically rigorous, and physically realistic discretization of the continuous theory in Chapter II.

sampled at a high enough resolution (i.e., if the grid in Figure 3(b) is fine enough) the integral of f over the continuous line is essentially equal to a weighted sum of the integer samples of f; see Figure 3(c). Moreover, the values of these weights follow patterns similar to our discrete "line" in Figure 2(b). The remainder of this chapter is devoted to making these notions rigorous and then exploring their consequences. As before, for the sake of readability we suppress the domains of the indices of summation; sums over the variables m, n, p, a, b, c range over $\mathbb{Z}_P$, while those over the variable q range over $\mathbb{Z}_Q$.

Specifically, we make the simplifying assumption that our continuous-variable periodic function $f : \mathbb{T}_P^3 \to \mathbb{R}$ has been sampled so finely that it is effectively

piecewise constant. That is, letting $\mathbf{f} \in \ell(\mathbb{Z}_P^3)$ denote the integer samples of f, we assume that for almost every $(x, y, z) \in \mathbb{T}_P^3$, $f(x, y, z)$ has the form

$$f(x,y,z) = \sum_{a,b,c \in \mathbb{Z}_P} \mathbf{f}[a,b,c] 1_{(a,b,c)+\left[-\frac{1}{2},\frac{1}{2}\right]^3}(x,y,z). \tag{27}$$

Here, the function $1_{(a,b,c)+\left[-\frac{1}{2},\frac{1}{2}\right]^3} : \mathbb{T}_P^3 \to \mathbb{R}$ is the characteristic function of the cube

$$\left\{(x,y,z) \in \mathbb{T}_P^3 : a - \frac{1}{2} \leq x \leq a + \frac{1}{2},\ b - \frac{1}{2} \leq y \leq b + \frac{1}{2},\ c - \frac{1}{2} \leq z \leq c + \frac{1}{2}\right\}.$$

Replacing (a, b, c) with (a', b', c') and substituting this expression for f into our continuous periodic operator (26) gives

$$(L_P f)(x,y,q) = \int_{\mathbb{T}_P} \sum_{a',b',c'} \mathbf{f}[a',b',c'] 1_{(a',b',c')+\left[-\frac{1}{2},\frac{1}{2}\right]^3}(x - z\psi_1[q], y - z\psi_2[q], z)\,\mathrm{d}z.$$

To proceed, we decompose the integral over $\mathbb{T}_P = \left[-\frac{1}{2}, P - \frac{1}{2}\right)$ into a sum of P integrals, each over an interval of the form $\left[p - \frac{1}{2}, p + \frac{1}{2}\right)$ for some $p = 0, ..., P - 1$:

$$(L_P f)(x,y,q) = \sum_p \int_{p-\frac{1}{2}}^{p+\frac{1}{2}} \sum_{a',b',c'} \mathbf{f}[a',b',c'] 1_{(a',b',c')+\left[-\frac{1}{2},\frac{1}{2}\right]^3}(x - z\psi_1[q], y - z\psi_2[q], z)\,\mathrm{d}z.$$

Distributing this integral over finite sums and scalar multiples gives

$$(L_P f)(x,y,q) = \sum_p \sum_{a',b',c'} \mathbf{f}[a',b',c'] \int_{p-\frac{1}{2}}^{p+\frac{1}{2}} 1_{(a',b',c')+\left[-\frac{1}{2},\frac{1}{2}\right]^3}(x - z\psi_1[q], y - z\psi_2[q], z)\,\mathrm{d}z.$$

At this point, we further assume that (x, y) lies on the integer grid. That is, we assume that $(x, y) = (m, n)$ for some $m, n \in \mathbb{Z}$. Thus, for any p, we can substitute

$(x - p\psi_1[q] - a, y - p\psi_2[q] - b, p - c)$ for (a', b', c'), yielding

$$\begin{aligned}
&(L_P f)(x, y, q) \\
&\quad = \sum_p \sum_{a,b,c} \mathbf{f}[x - p\psi_1[q] - a, y - p\psi_1[q] - b, p - c] \\
&\qquad \times \int_{p-\frac{1}{2}}^{p+\frac{1}{2}} 1_{(x-p\psi_1[q]-a, y-p\psi_2[q]-b, p-c)+\left[-\frac{1}{2},\frac{1}{2}\right]^3}(x - z\psi_1[q], y - z\psi_2[q], z)\,\mathrm{d}z \\
&\quad = \sum_p \sum_{a,b,c} \mathbf{w}_q[a, b, c]\mathbf{f}[x - p\psi_1[q] - a, y - p\psi_1[q] - b, p - c],
\end{aligned}$$

where $\mathbf{w}_q \in \ell(\mathbb{Z}_P^3)$ is the q-dependent weight function

$$\mathbf{w}_q[a, b, c] := \int_{p-\frac{1}{2}}^{p+\frac{1}{2}} 1_{(x-p\psi_1[q]-a, y-p\psi_2[q]-b, p-c)+\left[-\frac{1}{2},\frac{1}{2}\right]^3}(x - z\psi_1[q], y - z\psi_2[q], z)\,\mathrm{d}z.$$

We now simplify our expression for $\mathbf{w}_q$. First, by writing out the formula for the characteristic function and simplifying, we observe that

$$\mathbf{w}_q[a, b, c] = \int_{p-\frac{1}{2}}^{p+\frac{1}{2}} 1_{(a,b,c)+\left[-\frac{1}{2},\frac{1}{2}\right]^3}((z - p)\psi_1[q], (z - p)\psi_2[q], p - z)\,\mathrm{d}z.$$

Next, for any fixed p we substitute $t = z - p$ to obtain

$$\mathbf{w}_q[a, b, c] = \int_{-\frac{1}{2}}^{\frac{1}{2}} 1_{(a,b,c)+\left[-\frac{1}{2},\frac{1}{2}\right]^3}(t\psi_1[q], t\psi_2[q], -t)\,\mathrm{d}t. \tag{28}$$

Here, as t ranges from $-\frac{1}{2}$ to $\frac{1}{2}$, $(t\psi_1[q], t\psi_2[q], -t)$ traces out a segment of the line that, for a given q, passes through $(0, 0, 0)$ and $(\psi_1[q], \psi_2[q], -1)$. This line corresponds to the same discrete "line" that we sum over in our discrete CT model (18). Examining (28), we further note that since $-\frac{1}{2} \leq t \leq \frac{1}{2}$, in order for $\mathbf{w}_q(a, b, c)$ to

be non-zero, we must have that $c = 0$. Therefore, we can write (28) as a function of only two variables:

$$\mathbf{w}_q[a,b] = \int_{-\frac{1}{2}}^{\frac{1}{2}} 1_{(a,b)+\left[-\frac{1}{2},\frac{1}{2}\right]^2}(t\psi_1[q], t\psi_2[q])\,\mathrm{d}t. \tag{29}$$

We now summarize these results in the following theorem:

Theorem 2. *Making the assumption that the hyperspectral data cube $f(x,y,z)$ is a piecewise constant* (27)*, the continuous periodic X-ray transform* (26) *of $f : \mathbb{T}_P^3 \to \mathbb{R}$ at any $(x,y) = (m,n) \in \mathbb{Z}^2$ is a sum of the form*

$$(L_P f)(m,n,q) = \sum_{p\in\mathbb{Z}_P} \sum_{a,b\in\mathbb{Z}_P} \mathbf{w}_q[a,b]\mathbf{f}[m - p\psi_1[q] - a, n - p\psi_2[q] - b, p] \tag{30}$$

where $\mathbf{w}_q[a,b]$ is given by (29)*.*

Inspired by this result, given any $\psi_1, \psi_2 \in \ell(\mathbb{Z}_Q, \mathbb{Z})$ and weight functions $\{\mathbf{w}_q\}_{Q\in\mathbb{Z}_Q} \subseteq \ell(\mathbb{Z}_P^2)$, consider the operator $\mathbf{K} : \ell(\mathbb{Z}_P^3) \to \ell(\mathbb{Z}_P^2 \times \mathbb{Z}_Q)$,

$$(\mathbf{Kf})[m,n,q] = \sum_{p\in\mathbb{Z}_P} \sum_{a,b\in\mathbb{Z}_P} \mathbf{w}_q[a,b]\mathbf{f}[m - p\psi_1[q] - a, n - p\psi_2[q] - b, p]. \tag{31}$$

Theorem 2 tells us that the continuous X-ray transform of $f \in L^2(\mathbb{T}_P^3)$ can be well-approximated by sums of the form $\mathbf{Kf}$ where $\mathbf{f} \in \ell(\mathbb{Z}_P^3)$, provided the weights are taken according to (29). Moreover, note that our formula (31) for $\mathbf{K}$ is only a slight generalization of the formula (18) of our purely discrete X-ray transform $\mathbf{L}$. As such, it is reasonable to hope that the theory that we developed in Chapter III will carry over to this setting. We now show that this is indeed the case.

In the derivations that follow, it is helpful to regard the two-dimensional function $\mathbf{w}_q \in \ell(\mathbb{Z}_P^3)$ as a three-dimensional function $\mathbf{w}_q \in \ell(\mathbb{Z}_P^3)$ where $\mathbf{w}_q[a,b,c] := \mathbf{w}_q[a,b]\boldsymbol{\delta}_0[c]$. Note this implies $\mathbf{w}_q[a,b] = \sum_c \mathbf{w}_q[a,b,c]$, at which point our formula

for $\mathbf{K}$ becomes

$$(\mathbf{Kf})[m,n,q] = \sum_p \sum_{a,b,c} \mathbf{w}_q[a,b,c]\mathbf{f}[m - p\psi_1[q] - a, n - p\psi_2[q] - b, p - c].$$

As in Chapter III, it is not hard to show that this $\mathbf{K}$ is a well-defined bounded linear operator from $\ell(\mathbb{Z}_P^3)$ into $\ell(\mathbb{Z}_P^2 \times \mathbb{Z}_Q)$. Moreover, our method for computing $\mathbf{L}^*$ generalizes to a method for computing $\mathbf{K}^*$:

$$\begin{aligned}
&\langle \mathbf{f}, \mathbf{K}^*\mathbf{g}\rangle \\
&\quad = \langle \mathbf{Kf}, \mathbf{g}\rangle \\
&\quad = \sum_{m,n}\sum_q ((\mathbf{Kf})[m,n,q])^*\mathbf{g}[m,n,q] \\
&\quad = \sum_{a,b,c}\sum_q\sum_{m,n,p} (\mathbf{w}_q[a,b,c]\mathbf{f}[m - p\psi_1[q] - a, n - p\psi_2[q] - b, p - c])^*\mathbf{g}[m,n,q].
\end{aligned}$$

Continuing, for any $q \in \mathbb{Z}_Q$ and $a, b, c \in \mathbb{Z}_P$, we make the substitution $(m', n', p') = (m - p\psi_1[q] - a, n - p\psi_1[q] - b, p - c)$. Our expression for $\langle \mathbf{f}, \mathbf{K}^*\mathbf{g}\rangle$ then becomes

$$\begin{aligned}
&\sum_{a,b,c}\sum_q \sum_{m',n',p'} (\mathbf{f}[m',n',p'])^*(\mathbf{w}_q[a,b,c])^* \\
&\qquad\qquad \times \mathbf{g}[m' + (p'+c)\psi_1[q] + a, n' + (p'+c)\psi_2[q] + b, q] \\
&\quad = \sum_{m',n',p'} (\mathbf{f}[m',n',p'])^* \\
&\qquad \times \sum_q\sum_{a,b,c} (\mathbf{w}_q[a,b,c])^*\mathbf{g}[m' + (p'+c)\psi_1[q] + a, n' + (p'+c)\psi_2[q] + b, q].
\end{aligned}$$

At this point, we replace (a, b, c) with its negative, and write $(\mathbf{w}_q[-a,-b,-c])^*$ as $\tilde{\mathbf{w}}_q[a,b,c]$; in the signal processing literature, the conjugate-reversal $\tilde{\mathbf{w}}_q$ of $\mathbf{w}$ is known as the *involution* of $\mathbf{w}$. This yields the following expression for $\langle \mathbf{f}, \mathbf{K}^*\mathbf{g}\rangle$:

$$\sum_{m',n',p'} (\mathbf{f}[m',n',p'])^* \sum_q\sum_{a,b,c} \tilde{\mathbf{w}}_q[a,b,c]\mathbf{g}[m' + (p'-c)\psi_1[q] - a, n' + (p'-c)\psi_2[q] - b, q].$$

Thus, the adjoint of $\mathbf{K}$ is necessarily

$$(\mathbf{K}^*\mathbf{g})[m,n,p] = \sum_q \sum_{a,b,c} \tilde{\mathbf{w}}_q[a,b,c]\mathbf{g}[m+(p-c)\psi_1[q]-a, n+(p-c)\psi_2[q]-b, q].$$

Since $\mathbf{w}_q[a,b,c] = \mathbf{w}_q[a,b]\boldsymbol{\delta}_0(c)$ and $\tilde{\mathbf{w}}_q[a,b,c] = \tilde{\mathbf{w}}_q[a,b]\boldsymbol{\delta}_0(c)$ this becomes

$$(\mathbf{K}^*\mathbf{g})[m,n,p] = \sum_{q\in\mathbb{Z}_Q} \sum_{a,b\in\mathbb{Z}_P} \tilde{\mathbf{w}}_q[a,b]\mathbf{g}[m+p\psi_1[q]-a, n+p\psi_2[q]-b, q]. \tag{32}$$

Note that this expression for $\mathbf{K}^*$ is a generalization of that for $\mathbf{L}^*$. Indeed, it becomes $\mathbf{L}^*$ provided we take $\mathbf{w}_q = \boldsymbol{\delta}_{(0,0)}$. Similarly, our expression (20) for $\mathbf{L}^*\mathbf{L}$ generalizes to one for $\mathbf{K}^*\mathbf{K}$:

$$\begin{aligned}
&(\mathbf{K}^*\mathbf{K}\mathbf{f})[m,n,p] \\
&= \sum_q \sum_{a,b,c} \tilde{\mathbf{w}}_q[a,b,c](\mathbf{K}\mathbf{f})[m+(p-c)\psi_1[q]-a, n+(p-c)\psi_2[q]-b, q] \\
&= \sum_q \sum_{a',b',c'} \sum_{a,b,c} \mathbf{w}_q[a',b',c']\tilde{\mathbf{w}}_q[a,b,c] \\
&\quad \times \sum_{p'} \mathbf{f}[m+(p-p'-c)\psi_1[q]-a-a', n+(p-p'-c)\psi_2[q]-b-b', p'-c'].
\end{aligned}$$

Making the substitution $r = p' - p + c$, our expression now becomes

$$\begin{aligned}
(\mathbf{K}^*\mathbf{K}\mathbf{f})[m,n,p] = &\sum_q \sum_{a',b',c'} \sum_{a,b,c} \mathbf{w}_q[a',b',c']\tilde{\mathbf{w}}_q[a,b,c] \\
&\times \sum_r \mathbf{f}[m-r\psi_1[q]-a-a', n-r\psi_2[q]-b-b', p+r-c-c'].
\end{aligned}$$

Recalling that $\mathbf{L}^*\mathbf{L}$ could be expressed as a convolution, it is not surprising that $\mathbf{K}^*\mathbf{K}$ follows a similar pattern. To see this, recall the functions $\{\mathbf{h}_q\}_{q\in\mathbb{Z}_Q}$ defined in (22) which correspond to "characteristic functions" of discrete "lines." The last summations in our above expression for $\mathbf{K}^*\mathbf{K}$ can be written in terms of convolutions

of $\mathbf{f}$ with these $\mathbf{h}_q$'s:

$$\begin{aligned}
&(\mathbf{K}^*\mathbf{K}\mathbf{f})[m,n,p] \\
&\quad = \sum_q \sum_{a',b',c'} \sum_{a,b,c} \mathbf{w}_q[a',b',c']\tilde{\mathbf{w}}_q[a,b,c](\mathbf{h}_q * \mathbf{f})[m-a-a', n-b-b', p-c-c'] \\
&\quad = \sum_q \sum_{a,b,c} \tilde{\mathbf{w}}_q[a,b,c][\mathbf{w}_q * (\mathbf{h}_q * \mathbf{f})][m-a, n-b, p-c] \\
&\quad = \sum_q \left(\tilde{\mathbf{w}}_q * [\mathbf{w}_q * (\mathbf{h}_q * \mathbf{f})]\right)[m,n,p].
\end{aligned}$$

Since convolutions are associative, commutative, and distributive, we can rewrite this expression for $\mathbf{K}^*\mathbf{K}$ as

$$(\mathbf{K}^*\mathbf{K}\mathbf{f}) = \left(\sum_{q\in\mathbb{Z}_Q} [(\mathbf{w}_q * \tilde{\mathbf{w}}_q) * \mathbf{h}_q]\right) * \mathbf{f}. \tag{33}$$

That is, $\mathbf{K}^*\mathbf{K}$ is the act of filtering by $\sum_q (\mathbf{w} * \tilde{\mathbf{w}}_q) * \mathbf{h}_q$. As a quick check on our derivation, note that (33) becomes (23) when we take $\mathbf{w}_q = \tilde{\mathbf{w}}_q = \boldsymbol{\delta}_{(0,0,0)}$ for all q.

Since $\mathbf{K}^*\mathbf{K}$ can be expressed as a convolution, in order to understand it better, we once again turn to the DFT. Recall from (25) that

$$\begin{aligned}
\boldsymbol{\lambda}_q[\alpha,\beta,\gamma] &= (\mathbf{E}^*\mathbf{h}_q)[\alpha,\beta,\gamma] \\
&= P \begin{cases} 1, & \gamma = \alpha\psi_1[q] + \beta\psi_2[q] \mod P, \\ 0, & \text{else.} \end{cases}
\end{aligned} \tag{34}$$

Next, it is well-known that the DFT of $\tilde{\mathbf{w}}$ is the complex conjugate of the DFT of $\mathbf{w}$. Indeed, replacing (a,b,c) with its negative gives

$$(\mathbf{E}^*\tilde{\mathbf{w}}_q)[\alpha,\beta,\gamma] = \left(\sum_{a,b,c} \mathrm{e}^{-\frac{2\pi\mathrm{i}}{P}(\alpha a' + \beta b' + \gamma c')} \mathbf{w}_q[a,b,c]\right)^* = \{(\mathbf{E}^*\mathbf{w}_q)[\alpha,\beta,\gamma]\}^*.$$

Now recall that these $\mathbf{w}_q$'s are intended to be the three-dimensional extensions of those two-dimensional weight functions defined in (29). That is, given $\mathbf{w}_q \in \ell(\mathbb{Z}_P^2)$, we regard it as $\mathbf{w}_q \in \ell(\mathbb{Z}_P^3)$ by letting $\mathbf{w}_q[a,b,c] := \mathbf{w}_q[a,b]\boldsymbol{\delta}_0[c]$. The three-dimensional DFTs of such functions are equal to the two-dimensional DFTs of their restriction to $\mathbb{Z}_P^2$:

$$\begin{aligned}(\mathbf{E}^*\mathbf{w}_q)[\alpha,\beta,\gamma] &= \sum_{a,b,c} \mathrm{e}^{-\frac{2\pi\mathrm{i}}{P}(\alpha a+\beta b+\gamma c)}\mathbf{w}_q[a,b,c] \\ &= \sum_{a,b,c} \mathrm{e}^{-\frac{2\pi\mathrm{i}}{P}(\alpha a+\beta b+\gamma c)}\mathbf{w}_q[a,b]\boldsymbol{\delta}_0[c] \\ &= \sum_{a,b} \mathrm{e}^{-\frac{2\pi\mathrm{i}}{P}(\alpha a+\beta b)}\mathbf{w}_q[a,b] \\ &= (\mathbf{E}^*\mathbf{w}_q)[\alpha,\beta].\end{aligned}$$

As such, we actually have

$$(\mathbf{E}^*\mathbf{K}^*\mathbf{K}\mathbf{f})[\alpha,\beta,\gamma] = \sum_{q\in\mathbb{Z}_Q} |(\mathbf{E}^*\mathbf{w}_q)[\alpha,\beta]|^2\, \boldsymbol{\lambda}_q[\alpha,\beta,\gamma](\mathbf{E}^*\mathbf{f})[\alpha,\beta,\gamma]. \tag{35}$$

We summarize (31), (32), (34), and (35) in the following theorem:

Theorem 3. *Given $\psi_1,\psi_2 \in \ell(\mathbb{Z}_Q,\mathbb{Z})$ and $\{\mathbf{w}_q\}_{q\in\mathbb{Z}_Q} \subseteq \ell(\mathbb{Z}_P^2)$, consider the corresponding weighted discrete X-ray transform $\mathbf{K}:\ell(\mathbb{Z}_P^3)\to\ell(\mathbb{Z}_P^2\times\mathbb{Z}_Q)$,*

$$(\mathbf{K}\mathbf{f})[m,n,q] = \sum_{p\in\mathbb{Z}_P}\sum_{a,b\in\mathbb{Z}_P} \mathbf{w}_q[a,b]\mathbf{f}[m-p\psi_1[q]-a, n-p\psi_2[q]-b, p].$$

The adjoint of $\mathbf{K}$ is $\mathbf{K}^*:\ell(\mathbb{Z}_P^2\times\mathbb{Z}_Q)\to\ell(\mathbb{Z}_P^3)$,

$$(\mathbf{K}^*\mathbf{g})[m,n,p] = \sum_{q\in\mathbb{Z}_Q}\sum_{a,b\in\mathbb{Z}_P} (\mathbf{w}_q[a,b])^*\mathbf{g}[m+p\psi_1[q]+a, n+p\psi_2[q]+b, q].$$

Moreover, $\mathbf{K}^*\mathbf{K} : \ell(\mathbb{Z}_P^3) \to \ell(\mathbb{Z}_P^3 \times \mathbb{Z}_Q)$ *is a filter. Specifically, for any* $\mathbf{f} \in \ell(\mathbb{Z}_P^3)$,

$$(\mathbf{E}^*\mathbf{K}^*\mathbf{K}\mathbf{f})[\alpha, \beta, \gamma] = \boldsymbol{\kappa}[\alpha, \beta, \gamma](\mathbf{E}^*\mathbf{f})[\alpha, \beta, \gamma],$$

where the eigenvalue function $\boldsymbol{\kappa} \in \ell(\mathbb{Z}_P^3)$ *is*

$$\boldsymbol{\kappa}[\alpha, \beta, \gamma] = P \sum_{q \in \mathbb{Z}_Q} |(\mathbf{E}^*\mathbf{w}_q)[\alpha, \beta]|^2 \begin{cases} 1, & \gamma = \alpha\psi_1[q] + \beta\psi_2[q] \mod P, \\ 0, & \textit{else.} \end{cases}$$

To see how this result helps us to better understand $\mathbf{K}^*\mathbf{K}$, we now examine $\boldsymbol{\kappa}$ in the example where ψ_1 and ψ_2 correspond to the "knight's moves" of (17), depicted in Figure 1. In particular, for $q = 0$, we have $\psi_1[0] = 2$ and $\psi_2[0] = 1$. In this case, our expression (29) for our zeroth weight function is

$$\mathbf{w}_0[a, b] = \int_{-\frac{1}{2}}^{\frac{1}{2}} 1_{(a,b)+\left[-\frac{1}{2},\frac{1}{2}\right]^2}(2t, t)\, \mathrm{d}t.$$

Note $\mathbf{w}_0[a, b] = 0$ unless $b = 0$. Moreover, for $b = 0$ it turns out that $\mathbf{w}_0[a, b] \neq 0$ only when $a = -1, 0, 1$. To be precise, in this case, it is easy to show that

$$\mathbf{w}_0[a, b] = \begin{cases} \frac{1}{2}, & (a, b) = (0, 0), \\ \frac{1}{4}, & (a, b) = (1, 0), \\ \frac{1}{4}, & (a, b) = (-1, 0), \\ 0, & \text{else.} \end{cases}$$

Meanwhile when $q = 1$, we have $\psi_1[1] = 1, \psi_2[1] = 2$, and so

$$\mathbf{w}_1[a,b] = \int_{-\frac{1}{2}}^{\frac{1}{2}} 1_{(a,b)+\left[-\frac{1}{2},\frac{1}{2}\right]^2}(t,2t)\,\mathrm{d}t.$$

The calculation of $\mathbf{w}_1$ yields a result similar to the one for $\mathbf{w}_0$:

$$\mathbf{w}_1[a,b] = \begin{cases} \frac{1}{2}, & (a,b) = (0,0), \\ \frac{1}{4}, & (a,b) = (0,1), \\ \frac{1}{4}, & (a,b) = (0,-1), \\ 0, & \text{else.} \end{cases}$$

Further calculations shows that $\mathbf{w}_3$, $\mathbf{w}_4$ and $\mathbf{w}_7$ are all equal to $\mathbf{w}_0$, while $\mathbf{w}_2$, $\mathbf{w}_5$ and $\mathbf{w}_6$ equal $\mathbf{w}_1$, reflecting the inherent symmetries of the "knight's moves" set. In order to understand the ramifications of Theorem 3 in this setting, we must compute the DFTs of the $\mathbf{w}_q$'s. In particular, the DFT of $\mathbf{w}_0$ is

$$\begin{aligned} (\mathbf{E}^*\mathbf{w}_0)[\alpha,\beta] &= \sum_{a,b} \mathrm{e}^{-\frac{2\pi\mathrm{i}}{P}(\alpha a+\beta b)}\mathbf{w}_0[a,b] \\ &= \frac{1}{2}\mathrm{e}^0 + \frac{1}{4}\mathrm{e}^{-\frac{2\pi\mathrm{i}}{P}\alpha} + \frac{1}{4}\mathrm{e}^{-\frac{2\pi\mathrm{i}}{P}(-\alpha)} \\ &= \frac{1}{2} + \frac{1}{2}\cos\left(\tfrac{2\pi\alpha}{P}\right) \\ &= \frac{1}{2}\left(1+\cos\left(\tfrac{2\pi\alpha}{P}\right)\right) \\ &= \cos^2\left(\tfrac{\pi\alpha}{P}\right). \end{aligned}$$

In a similar manner, it can be shown that the DFT of $\mathbf{w}_1$ (which equals that of $\mathbf{w}_2$,$\mathbf{w}_5$, and $\mathbf{w}_6$) is $\cos^2(\pi\beta/P)$. Thus, for the "knight's moves" example, $\mathbf{K}^*\mathbf{K}$ acts

as a pointwise multiplication operator in the Fourier domain, multiplying $\mathbf{E}^*\mathbf{f}$ by $\boldsymbol{\kappa}$ which, by Theorem 3, is given by

$$\begin{aligned}\boldsymbol{\kappa}[\alpha,\beta,\gamma] &= \sum_q \left|(\mathbf{E}^*\mathbf{w}_q)[\alpha,\beta]\right|^2 \boldsymbol{\lambda}_q[\alpha,\beta,\gamma] \\ &= P\cos^4(\tfrac{\pi\alpha}{P}) \sum_{q\in\{0,3,4,7\}} \begin{cases} 1, & \gamma = \alpha\psi_1[q] + \beta\psi_2[q] \mod P, \\ 0, & \text{else} \end{cases} \\ &\quad + P\cos^4(\tfrac{\pi\beta}{P}) \sum_{q\in\{1,2,5,6\}} \begin{cases} 1, & \gamma = \alpha\psi_1[q] + \beta\psi_2[q] \mod P, \\ 0, & \text{else.} \end{cases}\end{aligned}$$

We now compare this result—obtained via the more realistic CT model introduced in this chapter—against that given in Chapter III. There, the unweighted discrete operator $\mathbf{L}^*\mathbf{L}$ also acts a pointwise multiplication operator in the Fourier domain. But whereas $\mathbf{E}^*\mathbf{K}^*\mathbf{K}\mathbf{f} = \boldsymbol{\kappa}\mathbf{E}^*\mathbf{f}$ where $\boldsymbol{\kappa}$ is given above, $\mathbf{L}^*\mathbf{L}$ instead satisfies $\mathbf{E}^*\mathbf{L}^*\mathbf{L}\mathbf{f} = \boldsymbol{\lambda}\mathbf{E}^*\mathbf{f}$, where $\boldsymbol{\lambda}$ is an unweighted version of $\boldsymbol{\kappa}$, namely

$$\boldsymbol{\lambda}[\alpha,\beta,\gamma] = \sum_{q=0}^{7} \begin{cases} 1, & \gamma = \alpha\psi_1[q] + \beta\psi_2[q] \mod P, \\ 0, & \text{else.} \end{cases}$$

That is, the main difference between the two approaches is the presence of the $\cos^4(\frac{\pi\alpha}{P})$ and $\cos^4(\frac{\pi\beta}{P})$ weights in the more realistic model—more realistic in the sense that $\mathbf{K}$ more closely imitates the continuous X-ray transform than $\mathbf{L}$ does. These weights in frequency space are nonnegative, no more than 1 in magnitude, and decay to 0 at the edges of our $\left[-\frac{P}{2},\frac{P}{2}\right]^2$ grid. This decay means that $\mathbf{K}^*\mathbf{K}$ does a worse job of preserving $\mathbf{f}$'s high frequency content than $\mathbf{L}^*\mathbf{L}$ does. This makes sense: by design, $\mathbf{K}^*\mathbf{K}$ does a better job modeling the continuous "blur" induced by our prism. Such blurring will, in reality, destroy high spatial frequency content.

We thus see that we pay a price for making our discrete model more faithful to the continuous one: given $\mathbf{L}^*\mathbf{L}\mathbf{f}$, we can recover all parts of the DFT of $\mathbf{f}$ that lie on the planes perpendicular to $\{(\psi_1[q], \psi_2[q], -1)\}_{q\in\mathbb{Z}_Q}$; given $\mathbf{K}^*\mathbf{K}\mathbf{f}$, the information contained in the edges of these planes is unreliable.

We emphasize that the above observations regarding $\mathbf{L}^*\mathbf{L}$ versus $\mathbf{K}^*\mathbf{K}$ were made in the special case where $\{(\psi_1[q], \psi_2[q]\}_{q\in\mathbb{Z}_Q}$ corresponded to the knight's moves of (17). We do not know whether this same behavior occurs in general. Nevertheless, there are some things we can say. First, both the operator $\mathbf{L}^*\mathbf{L}$ of Theorem 1 and $\mathbf{K}^*\mathbf{K}$ of Theorem 3 are filters, and the only difference between their eigenvalue functions $\boldsymbol{\lambda}$ and $\boldsymbol{\kappa}$ is the presence of additional weight terms in $\boldsymbol{\kappa}$. More can be said when we require these weights $|(\mathbf{E}^*\mathbf{w}_q)[\alpha, \beta]|^2$ to arise from $\mathbf{w}_q$'s of the form (29), such as is needed for Theorem 3 to hold. In particular, such $\mathbf{w}_q$'s necessarily have nonnegative values. Moreover, like we saw in the above example, these values necessarily sum to 1:

$$\begin{aligned}
\sum_{a,b\in\mathbb{Z}_P} \mathbf{w}_q[a,b] &= \sum_{a,b\in\mathbb{Z}_P} \int_{-\frac{1}{2}}^{\frac{1}{2}} 1_{(a,b)+\left[-\frac{1}{2},\frac{1}{2}\right]^2}(t\psi_1[q], t\psi_2[q])\,\mathrm{d}t \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\sum_{a,b\in\mathbb{Z}_P} 1_{(a,b)+\left[-\frac{1}{2},\frac{1}{2}\right]^2} \right)(t\psi_1[q], t\psi_2[q])\,\mathrm{d}t \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} 1_{\left[-\frac{1}{2},P+\frac{1}{2}\right]^2}(t\psi_1[q], t\psi_2[q])\,\mathrm{d}t \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} 1\,\mathrm{d}t \\
&= 1.
\end{aligned}$$

This forces the corresponding weights in the frequency domain to be no more than 1 for every $(\alpha, \beta) \in \mathbb{Z}_P^2$:

$$\begin{aligned}
|(\mathbf{E}^*\mathbf{w}_q)[\alpha,\beta]|^2 &= \left|\sum_{a,b} \mathrm{e}^{-\frac{2\pi \mathrm{i}}{P}(\alpha a+\beta b)}\mathbf{w}_q[a,b]\right|^2 \\
&\leq \left(\sum_{a,b}\left|\mathrm{e}^{-\frac{2\pi \mathrm{i}}{P}(\alpha a+\beta b)}\mathbf{w}_q[a,b]\right|\right)^2 \\
&= \left(\sum_{a,b}\mathbf{w}_q[a,b]\right)^2 \\
&= 1.
\end{aligned}$$

In order to determine the rate of decay of $|(\mathbf{E}^*\mathbf{w}_q)[\alpha,\beta]|^2$, we would probably need to first find explicit expressions for the DFTs of weight functions that have the form of (29). Our preliminary investigation of this quantity leads us to believe that such a computation is nontrivial. As such, we leave it for future work.

Regardless of whether we use $\mathbf{L}^*\mathbf{L}$ or $\mathbf{K}^*\mathbf{K}$, we are still faced with solving a linear system with a large null space. Because of this, we now regularize our least squares problem with penalty terms that will promote those solutions to the normal equations which are either sparse or have sparse gradients.

V. Chromotomographic reconstruction via the Split Bregman Method

Throughout this thesis our efforts have been focused toward creating and understanding a discrete model of the X-ray transform. From a computational standpoint, a discrete model is necessary: the best reconstruction algorithms are discrete in nature, and make use of modern computational hardware. In the following discussion, we present the ideas in terms of the discrete X-ray transform $\mathbf{L}$ of Chapter III. However, the same ideas can easily be applied to its generalization $\mathbf{K}$ given in Chapter IV. At its core, the CT reconstruction problem involves computing the $\mathbf{f}$ that solves the equation $\mathbf{Lf} = \mathbf{g}$ for given camera measurements, $\mathbf{g}$. However, as mentioned in the introduction, if our measurements are noisy, that is $\mathbf{g} = \mathbf{Lf}_0 + \boldsymbol{\varepsilon}$, this equation may not even have a solution.

This issue motivates us to turn to the method of linear least squares. That is, finding the $\mathbf{f}$ that makes $\mathbf{Lf}$ as close to $\mathbf{g}$ as possible by solving $\arg\min_{\mathbf{f}} \|\mathbf{Lf} - \mathbf{g}\|_2^2$. Since $\mathbf{L}$ is a bounded linear operator, the standard Hilbert space theory of linear least squares tells us that we can solve this minimization problem by instead solving the normal equations, $\mathbf{L}^*\mathbf{Lf} = \mathbf{L}^*\mathbf{g}$. To do this, in the previous chapters we have rigorously derived expressions for $\mathbf{L}^*$ and $\mathbf{L}^*\mathbf{L}$. Further investigation gave us that $\mathbf{L}^*\mathbf{L}$ is a type of operator known as a filter that is best understood from a frequency-based perspective. In particular, after examining it in the Fourier domain, we find that $\mathbf{L}^*\mathbf{L}$ destroys information; see Chapter III. Specifically, $\mathbf{L}^*\mathbf{L}$ has a nontrivial null space consisting of those functions whose Fourier transforms are supported on the set-complements of certain discrete planes.

Due to the size of $\mathbf{L}^*\mathbf{L}$'s null space, there are an infinite number of solutions to the normal equations, meaning we must make some further assumptions about the $\mathbf{f}$'s we wish to reconstruct. The intended purpose of the CT camera system is to monitor transient events that take place in natural scenes. In particular, if the transient event is very bright relative to the surrounding background, such as

a fireball, it is reasonable to restrict $\mathbf{f}$ to be *sparse*, meaning it has few nonzero entries. In addition to being a natural model for such signals, this allows us to make use of the many recent developments in the fields of sparse signal processing and compressed sensing. To be precise, of the infinite number of solutions $\mathbf{f}$ to the normal equations, we would like to choose the one solution whose zero "norm" $\|\mathbf{f}\|_0$ is as small as possible. Here $\|\mathbf{f}\|_0$ denotes the number of nonzero entries of $\mathbf{f}$. That is, in order to find the sparsest solution to the normal equations, we want to solve

$$\arg\min_{\mathbf{f}} \|\mathbf{f}\|_0 \text{ subject to } \mathbf{L}^*\mathbf{L}\mathbf{f} = \mathbf{L}^*\mathbf{g}.$$

However, the zero "norm" is not well-behaved: it is not a norm, nor is it differentiable, nor is it convex. Instead we will take the now-standard approach of using the 1-norm as a proxy for the zero "norm," and attempt to solve

$$\arg\min_{\mathbf{f}} \|\mathbf{f}\|_1 \text{ subject to } \mathbf{L}^*\mathbf{L}\mathbf{f} = \mathbf{L}^*\mathbf{g}. \tag{36}$$

Here, the 1-norm of $\mathbf{f} \in \ell(\mathbb{Z}_P^3)$ is defined to be

$$\|\mathbf{f}\|_1 := \sum_{m\in\mathbb{Z}_P} \sum_{n\in\mathbb{Z}_P} \sum_{p\in\mathbb{Z}_P} |\mathbf{f}[m,n,p]| \,.$$

It turns out that due to the size of $\mathbf{L}$'s null space, finding a direct solution to (36) is still too computationally difficult. The standard way of getting around this issue is to regularize (36) by adding a term that penalizes large values of $\|\mathbf{L}\mathbf{f} - \mathbf{g}\|_2^2$. That is, to find a sparse $\mathbf{f}$ for which $\mathbf{L}\mathbf{f}$ is close to $\mathbf{g}$, we want to solve the unconstrained optimization problem

$$\arg\min_{\mathbf{f}} \|\mathbf{f}\|_1 + \frac{\mu}{2}\|\mathbf{L}\mathbf{f} - \mathbf{g}\|_2^2, \tag{37}$$

where μ is some experimentally chosen weight. This problem is itself highly nontrivial. In short, if our task was to simply solve $\arg\min_{\mathbf{f}} \|\mathbf{L}\mathbf{f} - \mathbf{g}\|_2^2$ we would be able to find (many) solutions using standard linear least squares. On the other hand,

if $\mathbf{L}$ was the identity, it is known that we can solve $\min_{\mathbf{f}} \|\mathbf{f}\|_1 + \frac{\mu}{2}\|\mathbf{f} - \mathbf{g}\|_2^2$ using a technique known as *shrinkage*, or *soft-thresholding*:

$$\mathbf{f} = \text{shrink}(\mathbf{g}, \tfrac{1}{\mu}) := \text{sign}(\mathbf{g}) \odot \max\{|\mathbf{g}| - \tfrac{1}{\mu}, 0\}.$$

Here the sign and max functions are evaluated coordinate-wise, and the $\odot$ denotes pointwise vector multiplication. Unfortunately, our problem does not fit into either of these "simple" frameworks, and we must use a more complicated approach. In short, it is the combination of minimizing $\|\mathbf{f}\|_1$ and the 2-norm of a *function* of $\mathbf{f}$ that makes finding a solution to (37) so difficult.

To make this hard optimization problem more tractable, we instead attempt to solve a modified version of (37) obtained by transforming it into a two-variable problem. That is, we attempt to solve

$$\arg\min_{\mathbf{f},\mathbf{d}} \|\mathbf{d}\|_1 + \frac{\mu}{2}\|\mathbf{L}\mathbf{f} - \mathbf{g}\|_2^2 \text{ subject to } \mathbf{d} = \mathbf{f}.$$

Of course, this is simply a restatement of (37) and is truly no easier to directly solve. Nevertheless, this two-variable formulation helps us think about this problem in a new way. In particular, we can attempt to enforce our $\mathbf{d} = \mathbf{f}$ constraint by adding a new penalty term, once again yielding an unconstrained optimization problem:

$$\arg\min_{\mathbf{f},\mathbf{d}} \|\mathbf{d}\|_1 + \frac{\mu}{2}\|\mathbf{L}\mathbf{f} - \mathbf{g}\|_2^2 + \frac{\nu}{2}\|\mathbf{d} - \mathbf{f}\|_2^2. \tag{38}$$

In practice, in order to ensure that our $\mathbf{L}\mathbf{f} = \mathbf{g}$ and $\mathbf{d} = \mathbf{f}$ constraints are adequately enforced, it is common to try to solve (38) for larger and larger values of μ and ν and use the "eye test" to find the best apparent solution. Notice though, that as μ and ν get very large, the 1-norm term becomes insignificant and we are essentially again solving an ill-conditioned problem. Putting aside this issue for the moment, note that for any fixed μ and ν, it is not obvious how to find an optimal $\mathbf{d}$ and $\mathbf{f}$. One

approach to solving (38) involves alternating minimizations, where we iteratively calculate new values of $\mathbf{f}$ and $\mathbf{d}$ according to:

$$\mathbf{f}^{k+1} = \arg\min_{\mathbf{f}} \frac{\mu}{2}\|\mathbf{L}\mathbf{f} - \mathbf{g}\|_2^2 + \frac{\nu}{2}\|\mathbf{d}^k - \mathbf{f}\|_2^2,$$
$$\mathbf{d}^{k+1} = \arg\min_{\mathbf{d}} \|\mathbf{d}\|_1 + \frac{\nu}{2}\|\mathbf{d} - \mathbf{f}^{k+1}\|_2^2.$$

The issue with this approach is that $\mathbf{d}^{k+1} = \text{shrink}(\mathbf{f}^{k+1}, 1/\nu)$ meaning that $\mathbf{d}$ and $\mathbf{f}$ will never be equal.

Fortunately, the *Split Bregman* method of [11] gives a way to circumvent both of these problems. In the Split Bregman method we, like above, alternate between finding the optimal $\mathbf{d}$ for a given $\mathbf{f}$ and finding the optimal $\mathbf{f}$ for a given $\mathbf{d}$. However, unlike the above algorithm, the Split Bregman method introduces extra "noise" parameters that, when taken into account, allow this alternating optimization approach to converge. In order to understand the Split Bregman method, we must first understand its predecessor, known as Bregman Iteration.

5.1 Bregman Iteration

Bregman iteration [5] provides an alternative method for minimizing a given convex functional $E(\mathbf{u})$ according to a constraint of the form $H(\mathbf{u}) = 0$. Here, H is assumed to be a differentiable convex functional whose minimum value is 0. In the context of the CT reconstruction, $\mathbf{u}$ will become our pair of variables $(\mathbf{d}, \mathbf{f})$, and $H(\mathbf{u})$ and $E(\mathbf{u})$ will be taken to be

$$E(\mathbf{u}) = E(\mathbf{d}, \mathbf{f}) = \|\mathbf{d}\|_1 + \frac{\mu}{2}\|\mathbf{L}\mathbf{f} - \mathbf{g}\|_2^2,$$
$$H(\mathbf{u}) = H(\mathbf{d}, \mathbf{f}) = \frac{1}{2}\|\mathbf{d} - \mathbf{f}\|_2^2. \tag{39}$$

As mentioned in the previous section, the traditional method for enforcing the constraint $H(\mathbf{u}) = 0$ is to add a penalty term to our objective function $E(\mathbf{u})$. That is

we attempt to solve

$$\arg\min_{\mathbf{u}} E(\mathbf{u}) + \nu H(\mathbf{u}) \tag{40}$$

for larger and larger values of ν. This however can lead to numerical instabilities for large values of ν. As a remedy to this problem, Bregman iteration takes a different approach. Specifically, it uses the concept of "Bregman distance," which is the "error" between the convex functional E and its supporting linear approximation, namely the function

$$D_E^{\mathbf{p}}(\mathbf{u}, \mathbf{v}) = E(\mathbf{u}) - E(\mathbf{v}) - \langle \mathbf{p}, \mathbf{u} - \mathbf{v} \rangle.$$

Here, $\mathbf{p}$ is a *subgradient* of E at $\mathbf{v}$, namely a member of the *subdifferential* set

$$\partial E(\mathbf{v}) := \{\mathbf{p} \in [\ell(\mathbb{Z}_P^3)]^2 : E(\mathbf{u}) \geq E(\mathbf{v}) + \langle \mathbf{p}, \mathbf{u} - \mathbf{v} \rangle, \forall \mathbf{u}\}.$$

To be clear, Bregman iteration does not require E to be differentiable but rather, only convex. At any point $\mathbf{v}$ where E is differentiable the only $\mathbf{p} \in \partial E$ is its gradient $\nabla E(\mathbf{v})$. This can be understood by considering the first-order Taylor approximation, or linearization, of E at $\mathbf{v}$: $E(\mathbf{u}) \approx E(\mathbf{v}) + \langle \nabla E(\mathbf{v}), \mathbf{u} - \mathbf{v} \rangle$. Because E is convex, this approximation is from below, that is, $E(\mathbf{u}) \geq E(\mathbf{v}) + \langle \nabla E(v), u - v \rangle$ for all $\mathbf{u}$ and so $\nabla E(\mathbf{v}) \in \partial E(\mathbf{v})$. At points $\mathbf{v}$ at which E is not differentiable, one can show that there nevertheless exist vectors $\mathbf{p}$ for which $E(\mathbf{u}) \geq E(\mathbf{v}) + \langle \mathbf{p}, \mathbf{u} - \mathbf{v} \rangle$ for all $\mathbf{u}$. As stated above, any such vector $\mathbf{p}$ is called a subgradient of E at $\mathbf{v}$, and the (nonempty) collection of all such subgradients is denoted $\partial E(\mathbf{v})$.

In Bregman iteration, rather than attempt to solve (40) over and over again for larger and larger values of ν, we instead fix ν, make an initial guess $\mathbf{u}^0$, and iteratively solve a version of (40) where our objective function $E(\mathbf{u})$ is replaced with

the Bregman distance from a given $\mathbf{u}^k$:

$$\mathbf{u}^{k+1} = \arg\min_{\mathbf{u}} D_E^{\mathbf{p}^k}(\mathbf{u}, \mathbf{u}^k) + \nu H(\mathbf{u}).$$

Note this requires us to find an explicit subgradient $\mathbf{p}^k$ for E at every given $\mathbf{u}^k$. In practice, this is accomplished by choosing $\mathbf{u}^0$ to be any minimizer of the unconstrained functional $E(\mathbf{u})$, which permits us to take our initial subgradient to be $\mathbf{p}^0 = 0$. To derive a formula that iteratively updates $\mathbf{p}$, remember that H is differentiable, and since $\mathbf{u}^{k+1}$ minimizes $D_E^{\mathbf{p}^k}(\mathbf{u}, \mathbf{u}^k) + \nu H(\mathbf{u})$, subdifferential theory informs us that

$$\begin{aligned} 0 &\in \partial \left[D_E^{\mathbf{p}^k}(\mathbf{u}, \mathbf{u}^k) + \nu H(\mathbf{u}) \right] |_{\mathbf{u}=\mathbf{u}^{k+1}} \\ &= \partial \left[E(\mathbf{u}) - E(\mathbf{u}^k) - \langle \mathbf{p}^k, \mathbf{u} - \mathbf{u}^k \rangle + \nu H(\mathbf{u}) \right] |_{\mathbf{u}=\mathbf{u}^{k+1}}. \end{aligned}$$

Since $-E(\mathbf{u}^k) - \langle \mathbf{p}^k, \mathbf{u} - \mathbf{u}^k \rangle + \nu H(\mathbf{u})$ is differentiable, this simplifies to

$$\begin{aligned} 0 &\in \partial E(\mathbf{u})|_{\mathbf{u}=\mathbf{u}^{k+1}} + \nabla \left[-E(\mathbf{u}^k) - \langle \mathbf{p}^k, \mathbf{u} - \mathbf{u}^k \rangle + \nu H(\mathbf{u}) \right] |_{\mathbf{u}=\mathbf{u}^{k+1}} \\ &= \partial E(\mathbf{u})|_{\mathbf{u}=\mathbf{u}^{k+1}} - \mathbf{p}^k + \nu(\nabla H)(\mathbf{u}^{k+1}). \end{aligned}$$

Since $\mathbf{p}^{k+1} \in \partial E(\mathbf{u}^{k+1})$ this tell us it suffices to take

$$\mathbf{p}^{k+1} := \mathbf{p}^k - \nu \nabla H(\mathbf{u}^{k+1}).$$

To summarize, traditional Bregman iteration is to let $\mathbf{u}^0$ be a minimum of $E(\mathbf{u})$, let $\mathbf{p}^0 = 0$, and to then repeatedly apply the following two-step algorithm:

$$\mathbf{u}^{k+1} := \arg\min_{\mathbf{u}} D_E^{\mathbf{p}^k}(\mathbf{u}, \mathbf{u}^k) + \nu H(\mathbf{u}), \tag{41}$$

$$\mathbf{p}^{k+1} := \mathbf{p}^k - \nu \nabla H(\mathbf{u}^{k+1}). \tag{42}$$

In [11], the authors focus on the special case where $H(\mathbf{u})$ is assumed to have the form $\frac{1}{2}\|A\mathbf{u}-\mathbf{b}\|_2^2$. Note that this assumption is indeed valid in our setting since we define $H(\mathbf{u})$ as $H(\mathbf{d},\mathbf{f}) = \|\mathbf{d}-\mathbf{f}\|_2^2$, meaning we can take

$$A = \begin{bmatrix} I & -I \end{bmatrix}, \ \mathbf{u} = \begin{bmatrix} \mathbf{d} \\ \mathbf{f} \end{bmatrix}, \text{ and } \mathbf{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The reason they make this assumption in [11] is that it simplifies the Bregman iteration steps given in (41) and (42). Specifically, since $\nabla\|A\mathbf{u}-\mathbf{b}\|_2^2 = 2A^T(A\mathbf{u}-\mathbf{b})$, (41) and (42) become

$$\begin{aligned}
\mathbf{u}^{k+1} &= \arg\min_{\mathbf{u}} D_E^{\mathbf{p}^k}(\mathbf{u},\mathbf{u}^k) + \nu H(\mathbf{u}) \\
&= \arg\min_{\mathbf{u}} E(\mathbf{u}) - E(\mathbf{u}^k) - \langle \mathbf{p}^k, \mathbf{u}-\mathbf{u}^k\rangle + \frac{\nu}{2}\|A\mathbf{u}-\mathbf{b}\|_2^2 \\
&= \arg\min_{\mathbf{u}} E(\mathbf{u}) - \langle \mathbf{p}^k, \mathbf{u}-\mathbf{u}^k\rangle + \frac{\nu}{2}\|A\mathbf{u}-\mathbf{b}\|_2^2, \qquad (43) \\
\mathbf{p}^{k+1} &= \mathbf{p}^k - \nu A^T(A\mathbf{u}^{k+1}-\mathbf{b}). \qquad (44)
\end{aligned}$$

Moreover, in [11] they use the fact that this two-step process has a surprisingly simpler formulation: letting $\mathbf{u}^0$ be a minimizer of $E(\mathbf{u})$, $\mathbf{p}^0 = 0$ and $\mathbf{b}^0 = \mathbf{b}$, it turns out that the repeated application of (43) and (44) is equivalent to the repeated application of

$$\begin{aligned}
\mathbf{u}^{k+1} &= \arg\min_{\mathbf{u}} E(\mathbf{u}) + \frac{\nu}{2}\|A\mathbf{u}-\mathbf{b}^k\|_2^2, \qquad (45) \\
\mathbf{b}^{k+1} &= \mathbf{b}^k + \mathbf{b} - A\mathbf{u}^{k+1}. \qquad (46)
\end{aligned}$$

Looking at (46) we can see the true inspiration behind Bregman iteration: for a fixed ν the value $\mathbf{u}^{k+1}$ given by (45) will not exactly satisfy the constraint $A\mathbf{u} = \mathbf{b}$; in (46) we add the "error" $\mathbf{b} - A\mathbf{u}^{k+1}$ back into $\mathbf{b}^k$, thereby forcing, in the next iteration, the optimization (45) to better compensate for the aspects of $\mathbf{b}$ that it is missing. In

the image denoising literature—where Bregman iteration was first popularized—this step is known as "adding the noise back in."

We now apply these ideas in the special case where we have

$$E(\mathbf{u}) = E(\mathbf{d}, \mathbf{f}) = \|\mathbf{d}\|_1 + \frac{\mu}{2}\|\mathbf{L}\mathbf{f} - \mathbf{g}\|_2^2, \tag{47}$$

$$H(\mathbf{u}) = H(\mathbf{d}, \mathbf{f}) = \left\| \begin{bmatrix} I & -I \end{bmatrix} \begin{bmatrix} \mathbf{d} \\ \mathbf{f} \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\|_2^2. \tag{48}$$

Note that here, the initialization of picking $\mathbf{u}^0$ to be a minimum of $E(\mathbf{u})$ and $\mathbf{b}^0 = \mathbf{b}$ corresponds to taking $\mathbf{d}^0 = 0$, $\mathbf{f}^0$ to be any solution to $\mathbf{L}^*\mathbf{L}\mathbf{f} = \mathbf{L}^*\mathbf{g}$ and $\mathbf{b}^0 = 0$.

5.2 Split Bregman Iteration

We now turn to the recently introduced, state-of-the art numerical optimization technique known as the Split Bregman method in order to solve

$$\arg\min_{\mathbf{d},\mathbf{f}} \|\mathbf{d}\|_1 + \frac{\mu}{2}\|\mathbf{L}\mathbf{f} - \mathbf{g}\|_2^2 + \frac{\nu}{2}\|\mathbf{d} - \mathbf{f}\|_2^2.$$

Letting $\mathbf{d}^0 = 0$, $\mathbf{f}^0$ be any solution to $\mathbf{L}^*\mathbf{L}\mathbf{f} = \mathbf{L}^*\mathbf{g}$, and $\mathbf{b}^0 = 0$, applying the Bregman iteration steps (45) and (46) to the functionals given in (47) and (48) gives

$$(\mathbf{d}^{k+1}, \mathbf{f}^{k+1}) = \arg\min_{\mathbf{d},\mathbf{f}} \|\mathbf{d}\|_1 + \frac{\mu}{2}\|\mathbf{L}\mathbf{f} - \mathbf{g}\|_2^2 + \frac{\nu}{2}\|\mathbf{d} - \mathbf{f} - \mathbf{b}^k\|_2^2, \tag{49}$$

$$\mathbf{b}^{k+1} = \mathbf{b}^k + \mathbf{b} - (\mathbf{d}^{k+1} - \mathbf{f}^{k+1}). \tag{50}$$

This is Bregman iteration in this particular setting. The *Split*-Bregman method involves splitting the numerically difficult optimization (49) into two steps:

$$\mathbf{f}^{k+1} = \arg\min_{\mathbf{f}} \frac{\mu}{2}\|\mathbf{L}\mathbf{f} - \mathbf{g}\|_2^2 + \frac{\nu}{2}\|\mathbf{d}^k - \mathbf{f} - \mathbf{b}^k\|_2^2, \tag{51}$$

$$\mathbf{d}^{k+1} = \arg\min_{\mathbf{d}} \|\mathbf{d}\|_1 + \frac{\nu}{2}\|\mathbf{d} - \mathbf{f}^{k+1} - \mathbf{b}^k\|_2^2, \tag{52}$$

$$\mathbf{b}^{k+1} = \mathbf{b}^k + \mathbf{b} - (\mathbf{d}^{k+1} - \mathbf{f}^{k+1}). \tag{53}$$

We now discuss the optimizations (51) and (52) in greater detail. To be precise, (51) can be solved using standard linear least squares:

$$\mathbf{f}^{k+1} = \arg\min_{\mathbf{f}} \left|\left| \begin{bmatrix} \sqrt{\mu}\mathbf{L} \\ \sqrt{\nu}\mathrm{I} \end{bmatrix} [\mathbf{f}] - \begin{bmatrix} \sqrt{\mu}\mathbf{g} \\ \sqrt{\nu}(\mathbf{d}^k - \mathbf{b}^k) \end{bmatrix} \right|\right|_2^2.$$

This is equivalent to solving the corresponding set of normal equations:

$$(\mu\mathbf{L}^*\mathbf{L} + \nu\mathrm{I})\mathbf{f}^{k+1} = \mu\mathbf{L}^*\mathbf{g} + \nu\mathrm{I}(\mathbf{d}^k - \mathbf{b}^k).$$

Namely, we let $\mathbf{f}^{k+1} = (\mu\mathbf{L}^*\mathbf{L} + \nu\mathrm{I})^{-1}[\mu\mathbf{L}^*\mathbf{g} + \nu\mathrm{I}(\mathbf{d}^k - \mathbf{b}^k)]$. Here, $\mu\mathbf{L}^*\mathbf{L} + \nu\mathrm{I}$ is a filter meaning we can use the DFT to invert it. To be precise, in Chapter III, we showed how to use the DFT to unitarily diagonalize $\mathbf{L}^*\mathbf{L}$ in terms of its eigenvalues $\boldsymbol{\lambda}[\alpha, \beta, \gamma]$. The eigenvalues of $\mu\mathbf{L}^*\mathbf{L} + \nu\mathrm{I}$ are given by the function $\mu\boldsymbol{\lambda}[\alpha, \beta, \gamma] + \nu$. The second step (52) of our Split Bregman iteration can be solved using shrinkage:

$$\begin{aligned} \mathbf{d}^{k+1} &= \mathrm{shrink}(\mathbf{f}^{k+1} + \mathbf{b}^k, \frac{1}{\nu}) \\ &= \mathrm{sign}(\mathbf{f}^{k+1} + \mathbf{b}^k) \odot \max\{|\mathbf{f}^{k+1} + \mathbf{b}^k| - \frac{1}{\nu}, 0\}. \end{aligned}$$

To be clear, Split Bregman iteration truly calls for new $\mathbf{f}$'s and $\mathbf{d}$'s to be calculated until the difference between $\mathbf{f}$ and $\mathbf{d}$ falls within a pre-determined threshold, at which point a new $\mathbf{b}$ is calculated and the process is repeated. That is, one is truly

supposed to alternate between (51) and (52) many times before updating $\mathbf{b}$ in (53). However, the authors of [11] indicate that this method is not desirable. They instead advocate calculating $\mathbf{f}$, $\mathbf{d}$, and $\mathbf{b}$ consecutively, claiming that for many applications, optimal efficiency is obtained with only one iteration of the "inner loop." This is because the advantage gained by a slight refinement of $\mathbf{f}$ and $\mathbf{d}$ pales in comparison against the benefit of updating $\mathbf{b}$. We now summarize this algorithm:

Split Bregman Algorithm for discrete chromotomographic reconstruction:

While $\|\mathbf{f}^k - \mathbf{f}^{k-1}\|_2 >$ tolerance

$$\mathbf{f}^{k+1} = \arg\min_{\mathbf{f}} \frac{\mu}{2}\|\mathbf{Lf} - \mathbf{g}\|_2^2 + \frac{\nu}{2}\|\mathbf{d}^k - \mathbf{f} - \mathbf{b}^k\|_2^2$$

$$\mathbf{d}^{k+1} = \arg\min_{\mathbf{d}} \|\mathbf{d}\|_1 + \frac{\nu}{2}\|\mathbf{d} - \mathbf{f}^{k+1} - \mathbf{b}^k\|_2^2$$

$$\mathbf{b}^{k+1} = \mathbf{b}^k + \mathbf{b} - (\mathbf{d}^{k+1} - \mathbf{f}^{k+1})$$

end

5.3 Numerical Experimentation

In Chapter IV we introduced a generalization $\mathbf{K}$ of our discrete X-ray transform $\mathbf{L}$ that was more faithful to the underlying real-world physics of CT cameras. At the end of that chapter, we discussed an example which showed that the price of this realism is a larger null space. In short, the operator $\mathbf{K}$ is not as well-behaved as $\mathbf{L}$, and so we expect the problem of reconstructing $\mathbf{f}$ from $\mathbf{Kf}$ to be harder than that of reconstructing it from $\mathbf{Lf}$. This begs two questions. First, how much of a price—from the standpoint of numerical reconstruction error—do we actually pay for using a more realistic model? Second, does the numerical experimentation support our hope that real-world CT reconstruction is actually possible?

In this section, we provide some preliminary numerical experimentation that answers both of these questions. We find that reconstructing from $\mathbf{Kf}$ is indeed

less reliable than reconstructing from $\mathbf{Lf}$. However, the two approaches perform similarly, meaning that the theory of Chapter IV indeed delivers greater realism at little cost. More importantly, our simulations reveal that when using either $\mathbf{L}$ or $\mathbf{K}$, it is indeed feasible to reconstruct a sufficiently sparse hyperspectral image $\mathbf{f}$ from a very small number of CT camera measurements.

To be precise, we compare $\mathbf{L}$ and $\mathbf{K}$ by synthesizing a large number of sparse images $\mathbf{f}$ and then applying the Split Bregman reconstruction algorithm twice, once for $\mathbf{L}$ and once for $\mathbf{K}$. To be clear, we are interested in reconstructing $\mathbf{f}$'s that are sparse, that is images with very few bright spots. To test how well we can reconstruct $\mathbf{f}$ after taking measurements from $\mathbf{K}$ and $\mathbf{L}$, we create examples of sparse scenes, drawing inspiration from astronomical images. In this setting we expect a nearly black background with a few stars. To create this scene we pick a proportion of pixels that correspond to stars and incrementally increase this proportion to investigate how $\mathbf{L}$ and $\mathbf{K}$ operate on scenes that are less and less sparse. Due to the large null spaces of both $\mathbf{L}$ and $\mathbf{K}$, we expect both to perform poorly as $\mathbf{f}$ becomes nonsparse. For each given proportion of pixels, we randomly assign points in our $P \times P$ spatial grid where these stars will appear, and assume that at every star location, 10 percent of the star's spectrum will be nonzero. That is, these stars are sparse in the color spectrum as well. To be clear, in our simulated data, we allow our fixed proportion of star locations to overlap, just as two stars may align in our line of sight. After simultaneously applying the Split Bregman algorithm to the problem of reconstructing $\mathbf{f}$ from $\mathbf{Lf}$ and $\mathbf{Kf}$, we plot the relative error, $\|\mathbf{f} - \mathbf{f}_{\text{recon}}\|_2 / \|\mathbf{f}\|_2$, of the reconstructed image from the original as a function of the proportion of stars in the scene.

Looking at Figure 4, we first notice that as our hyperspectral image $\mathbf{f}$ becomes less and less sparse, this algorithm does a poor job of reconstructing the original image. This of course makes sense because $\mathbf{L}$ and $\mathbf{K}$ have large null spaces, and so it is unreasonable to expect to find $\mathbf{f}$ unless it is sparse. Indeed, the entire motivation behind the Split Bregman algorithm is the assumption of sparsity. The

graph also confirms, as we previously predicted, that we can better reconstruct the original image using data produced by **L** as opposed to the data produced by the more realistic operator **K**. Indeed, as detailed in Chapters III and IV, the frequency components of **f** that lie off of certain planes are annihilated by both operators. However, **K** also diminishes the high frequency components of **f** that lie on these planes. That is why, when looking at Figure 4, the more realistic model consistently underperforms the less realistic one. This is not a bad thing: no real-world system would truly perform as well as our **L** does; meanwhile, **K** may indeed do a good job modeling the underlying physics, and so its performance curve is probably a better indicator of when real-world CT reconstruction could be accomplished in practice. Regardless, the investigation of these operators is a step towards better understanding the CT reconstruction problem in all settings. We can use these models to better understand the limitations of current systems and to suggest improvements to future systems.

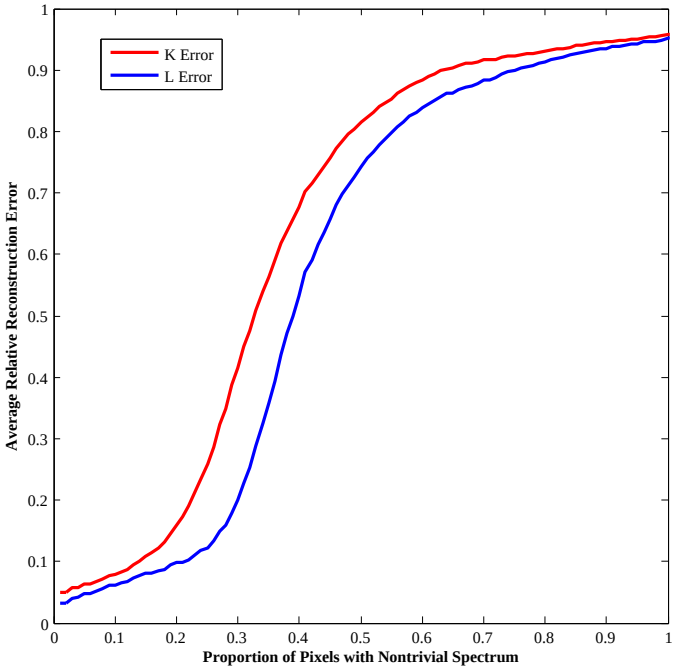

Figure 4: This plot shows the magnitudes of the relative errors between randomly generated $64 \times 64 \times 64$ simulated hyperspectral images and their reconstructions obtained via chromotomographic measurements. Specifically, it gives the magnitudes of these relative errors as a function of the proportion of stars in a simulated astronomical scene. For each proportion, 30 such images were randomly generated. Each star had an independently-generated 10%-sparse color spectrum. For each simulated image $\mathbf{f}$, we computed the corresponding chromotomographic camera measurements $\mathbf{g} = \mathbf{Lf}$ that arise from our discrete X-ray transform as well as those measurements $\mathbf{g} = \mathbf{Kf}$ that arise from its more-realistic cousin. In both cases, we used 8 discrete prism "angles" $\{(\psi_1[q], \psi_2[q])\}_{q=0}^{8}$. These corresponded to the "knight's moves" depicted in Figure 1. For each of these $\mathbf{g}$'s, we then used the Split Bregman algorithm to reconstruct $\mathbf{f}$ according to additional assumption that it is sparse. This is only computationally feasible due to the Fourier-based diagonalizations of $\mathbf{L}^*\mathbf{L}$ and $\mathbf{K}^*\mathbf{K}$ that we provided in Theorems 1 and 3, respectively; they let us perform the least-squares step of the Split Bregman algorithm using three-dimensional Fast Fourier Transforms. We draw two important conclusions from this graph. First, we see that chromotomographic reconstruction via the Split Bregman algorithm is indeed possible, and can be very successful provided the underlying image is sufficiently sparse. Second, we have numerical confirmation of the observation we made at the end of Chapter 4: it is indeed easier to reconstruct $\mathbf{f}$ from its idealized discrete X-ray transform measurements $\mathbf{Lf}$ than from its more realistic measurements $\mathbf{Kf}$. Based on its greater faithfulness to the underlying physics, we believe the higher of these two curves will be a better indication of the actual performance of a real-world chromotomographic imager, should such a system be fielded.

Bibliography

1. An, M., A. K. Brodzik, J. Mooney and R. Tolimieri. "Data restoration in chromotomographic hyperspectral imaging." *Proc. SPIE*, 4123:150–161, 2000.
2. Averbuch, A. and Y. Shkolnisky. "3D discrete X-ray transform." *Applied and Computational Harmonic Analysis*, 17:259–276, 2004.
3. Bostick, R. L. and G. P. Perram. "Hyperspectral imaging using chromotomography: A fieldable visible instrument for transient events." *International Journal of High Speed Electronics and Systems*, 18(3):519, 2008.
4. Bostick, R. and G. Perram. "Instrumental error in chromotomosynthetic hyperspectral imaging." *Applied Optics*, 51(21):5186–5200, 2012.
5. Bregman. L. "The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex optimization." *USSR Computational Mathematics and Mathematical Physics*, 7:200-217, 1967.
6. Brodzik, A. and J. Mooney. "Convex projections algorithm for restoration of limited-angle chromotographic images." *Journal of the Optical Society of America A. Optics, Image Science, and Vision*, 16(2):246–257, 1999.
7. Deming, R. W. "Chromotomography for a rotating-prism instrument using backprojection, then filtering." *Optics Letters*, 31(15):2281–2283, 2006.
8. Descour, M. and E. Dereniak. "Computed-tomography imaging spectrometer: experimental calibration and reconstruction results." *Applied Optics*, 34(22):4817–4826, 1995.
9. Frikel, J. "Sparse regularization in limited angle tomography." *Applied and Computational Harmonic Analysis*, 34(1):117–141, 2013.
10. Gehm, M. E., R. John, D. J. Brady, R. M. Willett and T. J. Schulz. "Single-shot compressive spectral imaging with a dual-disperser architecture." *Optical Society of America*, 15(21):14013–14027, 2007.
11. Goldstein, T. and S. Osher. "The Split Bregman Method for L1-Regularized Problems." *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.
12. Mooney, J., V. Vickers, M. An and A. Brodzik. "High-throughput hyperspectral infrared camera." *Optical Society of America*, 14(11):2951–2961, 1997.
13. O'Dell, D. *Development and demonstration of a field-deployable fast chromotomographic imager*. Master's Thesis, AFIT/GEO/ENP/10-M01, Department of Electrical and Computer Engineering, Air Force Institute of Technology, 2010.
14. Okamoto, T., A. Takahashi and I. Yamaguchi. "Simultaneous Acquisition of Spectral and Spatial Intensity Distribution." *Society for Applied Spectroscopy*, 47(8):1198–1202, 1993.

15. Osher, S., M. Burger, D. Goldfarb, J. Xu and W. Yin. "An iterative regularization method for total variation-based image restoration." *Multiscale Modeling and Simulation*, 4(2):460-489, 2005.

16. Starr, W. J., T. A. Book, A. Morse, S. D. Miller, E. D. Swenson, R. G. Cobb and J. T. Black. "Target Acquisition/Tracking for the Space-Based Chromotomographic Hyperspectral Imaging Experiment." *The American Institute for Aeronautics and Astronautics*, 7655–7675, 2010.

17. Su'e, Chad. *Characterization of a hyperspectral chromotomographic imaging ground system.* PhD dissertation, AFIT/EE.ABET/ENP/12-M03, Department of Electrical and Computer Engineering, Air Force Institute of Technology, 2012.

18. Wagadarikar, A. A., N. P. Pitsianis, X. Sun and D. J. Brady. "Video rate spectral imaging using a coded aperture snapshot spectral imager." *Optical Society of America*, 2009.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
21 Mar 2013	Master's thesis	Aug 2011–Mar 2013

4. TITLE AND SUBTITLE

A Discrete X-Ray Transform for Chromotomographic Hyperspectral Imaging

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Cooke, David J, 2d Lt

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (EN)
2950 Hobson Way
Wright-Patterson AFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENC-13-M-05

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Intentionally Left Blank

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The United States Air Force has a pressing need for new methods of hyperspectral imaging. All current hyperspectral imaging technologies require long exposure times, since each involves filtering the available light, either spatially or according to color. We consider a recently proposed method for hypserspectral imaging that promises shorter exposure times. This new method applies the mathematical principles of tomography to the hyperspectral data cube. Known as chromotomography, this method uses a spinning prism to essentially capture the integrals of this cube over many rotations of a single line. This thesis addresses some of the mathematical issues that arise when trying to reconstruct a hyperspectral image from chromotomographic measurements. After reviewing some of the mathematical shortcomings of the current state of the art—which arise from the technical difficulties of working with the continuous-variable X-ray transform—we make three contributions. First, we introduce a mathematically rigorous, discrete, X-ray transform that is somewhat faithful to its continuous cousin. Second, we show how under a few simplifying assumptions, our discrete transform can be generalized so as to provide a good approximation of the continuous one. This discretization allows us to apply modern finite-dimensional optimization methods to the chromotomographic reconstruction problem. Our third contribution is to apply a popular new example of such a method, known as Split Bregman iteration.

15. SUBJECT TERMS

Fourier transform, chromotomography, hyperspectral imaging

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Matthew C. Fickus (ENC)
U	U	U	UU	69	19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636 x 4513